Department of Molecular Biology
Faculty of Sciences
Universidad Autonoma de Madrid
Madrid, Spain

Functional characterization of the *Bacillus subtilis* conjugative plasmid pLS20: identification of relaxosome components and mechanism of transcriptional control of main conjugation promoter P_c

Doctoral thesis of Gayetri Ramachandran
Madrid, Spain 2014

Functional characterization of the *Bacillus subtilis* conjugative plasmid pLS20: identification of relaxosome components and mechanism of transcriptional control of main conjugation promoter P_c

Gayetri Ramachandran
Centro Biologia Molecular "Severo Ochoa"
Madrid, 2014

SUMMARY

Plasmid conjugation plays a significant role in the dissemination of antibiotic resistance and pathogenicity determinants. Understanding how conjugation is regulated is important to gain insights into these features. Little is known about regulation of conjugation systems present on plasmids from Gram-positive bacteria. The plasmid pLS20 is a native conjugative plasmid isolated from *Bacillus subtilis* "*natto*".

Our earlier laboratory results reported a global view of the regulatory system of the pLS20 conjugation genes. It was found that about 40 conjugation genes were regulated by a divergently oriented single gene rco_{LS20}, which encodes for the master regulator of conjugation. Rco_{LS20} is responsible for keeping the system in its default "OFF" state. Conjugation is activated by anti-repressor Rap_{LS20}, which belongs of the family of Rap proteins. However, the activity of Rap_{LS20} is regulated by a signaling peptide Phr^*_{LS20}. Even though, the key players involved in the regulation of the conjugation genes have been identified, our knowledge on regulation of the genetic switch responsible for activating conjugation is limited.

In the **first chapter** of the thesis, we have studied in detail the molecular mechanism regulating the pLS20 conjugation genes using different *in vivo* and *in vitro* approaches. At least three levels of regulation controls plasmid mediated-conjugation including overlapping and divergent promoter of differing strengths, maintaining the expression of the genes of conjugation and repression, along with a tri-functional repressor Rco_{LS20}, which regulates its own transcription and finally as a tetramer Rco_{LS20}, by DNA-looping. These complex regulatory levels permits the switching "OFF" and "ON" of the system of conjugation.

The **second chapter** of the thesis studies the formation of the relaxosome of plasmid pLS20. This nucleoprotein is essential during conjugation, permitting the transfer of DNA. In case of plasmid pLS20, we found that besides the relaxase, encoded by gene *58* of pLS20, and that we named Rel_{LS20}, the functional processing of *ori*T requires two additional proteins encoded by genes *56* and *57*.

INDEX

1. PRESENTACIÓN

INTRODUCCIÓN

El descubrimiento de los antibióticos permitió el tratamiento efectivo de las infecciones bacterianas. Sin embargo las bacterias pueden adquirir resistencia, y además este efecto se puede extender con rapidez. La propagación de bacterias resistentes es debida a la transferencia de ADN, que contiene genes de resistencia a antibióticos, mediante un proceso conocido como transferencia genética lateral **(/horizontal) (TGL)**. La TGL no sólo tiene importancia clínica, si no que también tiene repercusión en la industria y en la biotecnología. La TGL parece haber contribuido a la translocación de genes de virulencia, resistencia a antibióticos y de otros genes entre especies, que pueden no estar relacionados filogenéticamente. Además, análisis retrospectivos de genomas de bacterias sugieren que una parte sustancial de los genomas bacterianos tienen un origen externo y son adquiridos mediante TGL. La TGL incluye a diferentes agentes y consta de tres mecanismos principalmente, de los que la **conjugación** tiene el rango de hospedadores más amplio.

La conjugación está mediada por elementos conjugativos que se encuentran en plásmidos y por elementos integrativos conjugativos (ICEs). Los elementos conjugativos pueden ser fácilmente diseminados entre cepas, destacando la importancia de estudiar el mecanismo de conjugación. El proceso de conjugación y su regulación transcripcional han sido estudiados considerablemente en varios plásmidos presentes en bacterias Gram-negativas (G-). En comparación es poco lo que sabe sobre los sistemas de conjugación en bacterias Gram-positivas (G+), a pesar de que muchos de estos organismos como *Bacillus amylofaciens* o *B. anthracis* son de gran importancia para la industria o la medicina respectivamente.

Nuestro laboratorio se centra en la caracterización del plásmido conjugativo nativo pLS20, que fue inicialmente identificado en *Bacillus subtilis natto* (un organismo G+). Además de ser una bacteria de suelo, *B. subtilis* es un comensal intestinal en animales y humanos. Nuestro grupo ha desvelado los mecanismos que controlan la conjugación y hemos descubierto que los genes de conjugación del pLS20 son

activados en presencia de células receptoras libres del plásmido. Hemos identificado a la proteína represora Rco_{LS20} (un represor de la conjugación) que mantiene a la conjugación en estado de 'OFF', y al anti-represor Rap_{LS20} (un miembro de la familia Rap: Regulator Aspartate Phosphatase) que activa la conjugación. La actividad del anti-represor es inhibida por el péptido Phr_{LS20}, codificado por el pLS20, que es secretado por la célula y puede ser absorbido por otras células tras un procesamiento secundario. Finalmente, es el péptido señalizador Phr^*_{LS20} el que determina el momento en el que los genes de conjugación son activados. Además los genes *28-74* que constituyen el operón de la conjugación del plásmido fueron determinados mediante RNAseq. Aunque algunos aspectos del circuito regulador han sido resueltos, las bases moleculares de la regulación aún se desconocen.

De este modo, en el **primer capítulo** de esta tesis, los resultados describen el mecanismo de la regulación de la conjugación a nivel transcripcional. Se realizaron experimentos para identificar y localizar al promotor principal de los genes de conjugación, P_c, y al promotor del gen rco_{LS20} P_r así como la función de la proteína. Rco_{LS20} es una proteína de unión a ADN de tipo Hélice-Giro-Hélice, por lo que era importante determinar los sitios del operador a los que Rco_{L20} podría unirse y como podría reprimir la conjugación hasta que la maquinaria tuviera que ser cambiada al estado de 'ON'.

Los resultados obtenidos proporcionan evidencias de que la regulación de los genes de conjugación presentes en pLS20 está basada en un complejo interruptor genético. Su regulación consta de al menos tres capas para funcionar como un interruptor, que se enumeran a continuación: (i) los promotores del primer gen de la conjugación, P_c, y del represor, P_r, presentan distinta potencia y se orientan de manera divergente y solapada (ii) un Rco_{LS20} tri-funcional –un regulador transcripcional con capacidad de activar o reprimir a su propio promotor Pr a baja o elevada concentración respectivamente, mientras que reprime al fuerte y divergente promotor de conjugación Pc, y (iii) la represión de Pc, por un lazo de ADN mediada por la unión de Rco_{LS20} a los dos operadores separados por una corta distancia de

75pb. El efecto acumulativo de estas capas reprime fuertemente al promotor principal de la conjugación P_c cuando las condiciones para la conjugación no son propicias, y permite un rápido cambio al estado de ON en los genes de conjugación cuando se dan las condiciones apropiadas.

Las bases del mecanismo de conjugación entre plásmidos están conservadas. Para que un plásmido sea conjugativo, son necesarios una serie de genes que codifican para proteínas que (i) procesan el plásmido de ADN a la forma que va a ser transferida, que generalmente es ADN de doble cadena, y (ii) generan el tubo de conjugación asociado a la membrana, a través del que el ADNss es transportado. El tubo de conjugación intercelular es una forma del sistema de secreción de tipo IV. En la generación de la forma de ADNss plasmídico está implicada una relaxasa, que forma un complejo nucleoproteico llamado relaxosoma, que introduce un corte específico de sitio y de hebra dentro del origen de transferencia (*oriT*). La relaxasa permanece covalentemente unida al ADN cortado y el relaxosoma se asocia al canal conjugativo en formación por medio de la llamada proteína de acoplamiento. Tras la transferencia de la hebra de ADNss en la célula receptora por medio de la formación del canal de conjugación, la relaxasa asociada dirige la recircularización del ADNss en la célula receptora.

En el **segundo capítulo** de esta tesis, hemos estudiado la formación del llamado complejo de transferencia génica o relaxosoma del plásmido pLS20. Descubrimos que además de la relaxasa, codificada por el gen 58, que denominamos Rel_{LS20}, el procesamiento funcional del *ori*T necesita de los genes *56* and *57*. También descubrimos que la región oriT putativa del plásmido pLS20 se encuentra entre los genes 55 y 56.

Este trabajo nos ayuda a a) entender la regulación de la conjugación como el proceso energético que es y cómo varios niveles de regulación permiten controlar

eficientemente el estado ON y OFF del proceso b) identificar los componentes del complejo de regulación génica, un requisito previo a la conjugación.

OBJETIVO

Resultados iniciales de nuestro laboratorio identificaron a los actores principales (Rco_{LS20}, Rap_{LS20} y Phr^*_{LS20}) implicados en la regulación de la conjugación mediada por el plásmido pLS20, proporcionando una visión global del sistema. Sin embargo, los aspectos moleculares del sistema regulatorio deben ser entendidos en detalle, más aún, es necesario identificar a las proteínas que participan en la formación del complejo de transferencia génico, una etapa esencial de la conjugación. Los objetivos principales de esta tesis han sido enumerados a continuación:

1. Entender el cambio genético que regula la expresión de los genes de conjugación.

2. Elucidar los componentes del complejo de transferencia génica, seguido de la identificación de la región del origen de transferencia *ori*T del plásmido pLS20.

RESULTADOS Y DISCUSIÓN

Capítulo I:

Los resultados de nuestra RNAseq muestran que los genes *28-74* forman parte del operón de conjugación. De este modo, el gen 28 fue identificado como el primer gen del operón; su promotor fue llamado promotor de la conjugación, que parecía estar bajo el control de la proteína de unión a ADN Rco_{LS20}. Nuestro objetivo fue deducir cómo la proteína de unión a ADN Rco_{LS20} regula la conjugación transcripcionalmente.

En primer lugar intentamos dilucidar la posición del promotor de la conjugación P_c (fragmento I_c fusionado a *lac*Z) y descubrimos que se localizaba inusualmente lejos

del *gen 28*, a 5' del único sitio *Eco*RI encontrado en la región intergénica de los genes *rco$_{LS20}$-28*. Utilizando distintas técnicas, determinamos la secuencia como 5´-ttaaaaatttcactgaaatac-**TTtACA**-gttaaaaaaatgtc-TGt**TATctT**-3´, que es bastante similar al sitio de reconocimento del promotor putativo σ^A. A continuación mediante RNA seq y ensayos de extensión de cebador demostramos que estos resultados eran consistentes con nuestro análisis *in vivo*.

Seguidamente realizamos experimentos para averiguar como funciona la proteína de unión a ADN/regulador maestro de la conjugación Rco$_{LS20}$, mediante análisis *in vivo*, y descubrimos que se trata de una proteína tri-funcional. Las tres funciones incluyen la activación de su promotor P_r que se encuentra en la región intergénica entre *rco$_{LS20}$-28* (Fragmento I_r), a la vez que inhibe al promotor de la conjugación P_c a bajos niveles. A mayores niveles de inducción, reprime a su propio promotor P_r. Tras la inducción ectópica de Rco$_{LS20}$ (mediante un promotor inducible por IPTG) utilizando IPTG 50µM, observamos la mayor inducción de la represión del promotor P_r. Finalmente los resultados mostraron que Rco$_{LS20}$ es la única proteína del pLS20*cat* necesaria para la activación y represión de los promotores P_r y P_c respectivamente; a niveles mayores, reprime a su propio promotor mediante un retroalimentación negativa.

Mediante distintas técnicas, incluyendo RNAseq y ensayos de extensión de cebador, descubrimos un solapamiento entre el promotor de represión P_r y el promotor de conjugación P_c. Los resultados sugieren que hay una secuencia de un promotor dependiente de σ^A localizada a 5' del sitio de inicio de la transcripción del P_r, que fue identificada como 5´-aaGAtA- 17pb -TgTAAa-3`, solapando así con el promotor de la conjugación P_c. Además, cabe mencionar que estos promotores solapantes permiten a la ARN polimerasa actuar como un regulador transcripcional, ya que en un momento dado la polimerasa podría tener acceso a uno solo de los promotores.

Una vez identificada la proteína de unión a ADN que regula la función del promotor de la conjugación P_c, nos preguntamos cómo era capaz de hacerlo. Mediante la actividad β-galactosidasa de los promotores, observamos que una región de ~160pb que se localizaba 85pb por debajo del promotor P_c, contenía un operador de Rco_{LS20} que es necesario para la represión eficiente del P_c. Este sitio fue denominado Operador O_I. Resultados subsiguientes demostraron que hay otro operador de Rco_{LS20} en el fragmento III_c para el que utilizamos el término Operador O_{II}. Utilizamos una estrategia similar para analizar el efecto de Rco_{LS20} sobre la activación del promotor de la represión P_r. La región necesaria para la represión del promotor de la conjugación P_c es también necesaria para la activación del promotor de la represión P_r.

Los datos de los ensayos de retardo en gel (EMSA) mostraron que Rco_{LS20} se une a los fragmentos III y XII, mientras que no se une a los fragmentos X. Cabe destacar que los fragmentos III y XII contienen a los operadores O_{II} y O_I respectivamente, ya que Rco_{LS20} era capaz de unirse a estos fragmentos y de provocar un retraso en su movilidad. Derivados más pequeños de los fragmentos III y XII fueron denominados como los fragmentos IIIA y XIIA. Utilizando resultados previos y los programas de identificación de motivos MEME y BIOPROSPECTOR, buscamos motivos conservados en los dos operadores de Rco_{LS20}. Estos análisis revelaron la presencia del motivo conservado de 8pb CAGTGAAA que está presente siete veces en el O_{II} (fragmento IIIA) y cuatro veces en el Operador O_I (fragmento XIIA). Los motivos fueron nombrados de a1-a7 en el fragmento IIIA y de b1-b4 en el fragmento XIIA. La unión de Rco_{LS20} a los operadores O_I y O_{II} fue confirmada mediante ensayos de huella de ADN con DNasaI, en los que se demostró que Rco_{LS20} se une al par de promotores P_c/P_r y a otro sitio que se encuentra 75pb por debajo del promotor P_c.

Los dos operadores están separados por una distancia de 75 pb, basándonos en estos resultados podría sugerirse que la regulación eficiente del par de promotores P_c/P_r necesitaría de una estructura de tipo lazo de ADN. Pero una región tan pequeña

como 75pb, podría plantear un obstáculo, a no ser que posea una curva estática intrínseca. De este modo, realizamos un ensayo de permutación circular para identificar la curvatura del fragmento de ADN. Utilizamos tres fragmentos solapantes de 314pb con la curvatura predicha presente en distintas posiciones. Estos fragmentos fueron resueltos en un gel al 2% de poliacrilamida y agarosa (PAA) para visualizar la migración diferencial de los fragmentos de ADN. Los fragmentos migraron en el gel de agarosa de manera similar, mientras que cuando fueron resueltos el gel de PAA, la migración se basó en la posición de la curvatura. El fragmento con la curvatura estática en el medio corría más lentamente en el gel, probando la presencia de una curvatura estática en el fragmento. Las técnicas de ultracentrifugación analítica mostraron que el peso molecular medio calculado para la proteína era de 85.200±1.700, indicando que la proteína formaba un tetrámero en solución, que podría facilitar la formación de lazos de ADN por la proteína. Experimentos adicionales de desfase nos ayudaron a visualizar la regulación por formación de un lazo de ADN, ya que la inserción de 5 pb en la región intergénica provoca la pérdida de la represión mediada por Rco_{LS20}. El ADN con la inserción de 5 pb fue utilizado en un EMSA, sorprendentemente se observó que a concentraciones mayores de proteínas, los fragmentos de ADN no entraban en el gel, ya que el complejo ADN-proteína puede formar puentes intermoleculares que impiden la migración del complejo nucleoproteico en el gel. Estos resultados proporcionan evidencias de que el tetrámero de Rco_{LS20} media la regulación eficiente de la conjugación formando un lazo en el ADN.

Los genes de conjugación presentes en el plásmido pLS20 son regulados por un complejo interruptor genético, que comprende tres niveles de regulación: a) promotores solapados y orientados de manera divergente y con distinta potencia, b) autoactivación y represión del promotor débil por el regulador de la conjugación a bajas y altas concentraciones respectivamente, junto con la represión del promotor de la conjugación y c) por último, la generación de un lazo de ADN, confirmado por la unión del Rco_{LS20} sobre los dos operadores, lo que lleva a la formación de un corto

lazo. Cabe destacar la importancia de la combinación de los tres niveles, que permite una fuerte represión de la conjugación, manteniendo de este modo el sistema en un estado de OFF por defecto, y también un rápido cambio a una elevada inducción de la conjugación en su estado de ¨ON¨.

Capítulo II:

La segunda parte de esta tesis aborda la localización del origen de transferencia *ori*T y a los genes/proteínas que podrían estar implicados en la formación del complejo denominado relaxosoma. Mediante técnicas *in silico* e *in vivo*, descubrimos que las proteínas llamadas p56, p57 y p58 están implicadas en la formación del complejo nucleoproteico, codificado por los *genes 56*, *57* y *58* respectivamente. Descubrimos además que el *gen 58* codifica para una relaxasa, una proteína Mob que no es suficiente para catalizar la reacción, sino que requiere de las dos proteínas auxiliares p56 y p57. Los análisis *in silico* mostraron que p56 es una proteína tipo CopG, mientras que p57 podría contener un motivo tipo Ribbon-helix-helix, que es similar al de proteínas como TraY y TraM encontradas en el plásmido F.

El siguiente paso fue la identificación del origen de transferencia *ori*T del plásmido pLS20, para ello, clonamos fragmentos del plásmido pLS20*cat* en el plásmido no movilizable pUCTA2501. Se descubrió que la región intergénica entre los *genes 55* y *56* contiene el *ori*T del plásmido pLS20. Es posible proponer, basándonos en la similitud de las secuencias, que el sitio de corte de la región de transferencia del plásmido pLS20*cat* está dentro de la secuencia 5´-aatggtgccagtt-3´.

La formación del relaxosoma es el paso clave para la transferencia del ADN a las células receptoras, que marca la identificación del sitio *ori*T mediada por la relaxasa y otras proteínas accesorias, seguido de la actividad de corte de la hebra. Cobra así importancia identificar a las proteínas implicadas en la formación del sustrato de ADN necesario para la transferencia. En el caso del pLS20, descubrimos que Rel_{LS20}, pertenece a una nueva familia de proteínas mob llamadas MOB_{MG} de plásmidos

conjugativos, con un motivo conservado; W (X_4) H(X_2) T(X_3) HXH(X_4) E(X_4) R en el motivo III de los dominios de la proteína. La organización genética de los genes en la *región mob* de los seis plásmidos representativos de la familia $MOB_{(MG)}$ muestra que el gen de la relaxasa está precedido por dos genes, uno que codifica para una proteína tipo Cop y otro para una proteína tipo RHH. Así, podríamos haber descubierto un nuevo tipo de sistema procesador de ADN implicado en la conjugación, y formado por tres proteínas.

2.INTRODUCTION

I.1 What is Lateral Gene Transfer? (LGT)

Bacteria evolve rapidly, not just by mutations and rapid growth, but also by acquiring DNA through different processes collectively known as Lateral Gene Transfer (LGT) (see review (1)). LGT encompasses processes, which move and rearrange prokaryotic DNA. LGT occurs on a global scale hence, theoretically, any gene in any bacterium anywhere in any microbial biosphere might be mobilized and spread (for review see, (2, 3)).

LGT accounts for the widespread dispersion of antibiotic resistance genes among (pathogenic) bacteria. Bacterial resistance to antibiotics is a worldwide health problem (4). Nearly all-pathogenic bacteria have been reported to be resistant to multiple antibiotics, with vancomycin-resistant enterococci (VRE), methicillin-resistant *Staphylococcus aureus* (MRSA), multi-drug resistant (MDR) *Pseudomonas aeruginosa*, and MDR *Mycobacterium tuberculosis* being particularly notorious (4–6). To develop new strategies to combat drug-resistant bacteria it is necessary to understand the various aspects of LGT.

Besides the spread of antibiotic resistance genes among pathogenic bacteria, LGT plays also a critical role in the emergence of new pathogenic organisms by the dissemination of genes encoding virulence factors (e.g., toxins, adhesins, capsules, invasion properties, etc.) (7). The virulence gene that encodes for the diphtheria toxin is carried by a dormant β-prophage found in *Corynebacterium diphtheria* (8), while anthrax caused by *Bacillus anthracis*, is due to the toxin and capsule genes carried on two plasmids pXO1 and pXO2 present in the pathogen (9). Thus, virulence genes can be located on different mobile elements and the transfer of these elements to other cells can result in the appearance of new pathogenic strains. Moreover, virulence genes can also be located in "pathogenicity islands", which are embedded in the bacterial genomes (for review see (10)).

LGT is a major factor that shapes the genomic make up of bacterial species. Evidence shows that for *Escherichia coli* that out of 4,288 ORFs, 755 ORFs have been introduced into its genome from the *Salmonella* lineage in at least 234 lateral transfer events since their divergence (11).

The large impact has on antibiotic resistance, pathogenicity and evolutionary adaptation in general underline, the importance of understanding the mechanism by which prokaryotes acquire DNA horizontally. There are three main processes involved in LGT: a) transformation, b) transduction and c) conjugation. A detailed description of the three mechanisms is given below.

I.1.1 Transformation

The first evidence that LGT could occur was the recognition that virulence determinants could be transferred between pneumococci in infected mice by Griffith. This phenomenon was later shown to be mediated by a process called transformation (12). Transformation is the uptake of foreign DNA from the environment by cells, which is generally followed by stable integration of the absorbed DNA in the bacterial genome via homologous recombination. Bacteria are the only organisms with the ability to undergo transformation and is a genuine bacterial DNA transfer process, as the other mechanisms like conjugation and transduction depend on the genes found on mobile genetic elements (see **section I.2**) (for review, see (1, 3, 13)).

Almost all environmental systems contain extracellular DNA, the majority of DNA is derived from dead cells or broken viral particles, but it has been shown that cells can also actively secrete DNA into the environment (14). Some bacterial cells develop a regulated physiological state of *competence*, which involves many proteins to take up DNA and to integrate it into the bacterial genome (13). Competence is triggered in response to specific environmental conditions such as nutrient availability, quorum sensing and starvation (15). Development of competence involves the regulated expression of genes encoding for proteins that constitute the sophisticated DNA

uptake system, which is related to Type IV pilus formation and Type II and III secretion systems (16). With the exception of plasmids that reconstitute themselves in the new host, all genetic materials undergo homologous recombination, based on the sequence similarity between the two DNA molecules (3).

The ability to take up DNA by competence has been demonstrated for various bacteria like *Bacillus subtilis*, cyanobacteria, green sulphur bacteria and human pathogens like *Helicobacter*, *Neisseria*, *Pseudomonas* and *Streptococcus* (3, 17).

Figure I.1 summarizes the steps involved in transformation. Extracellular DNA is recognized by receptor proteins on the outer surface, while another complex of membrane proteins mediate the physical transport of the DNA across the cell envelope. During the transport across the envelope one strand is degraded and single-stranded DNA (ssDNA) molecule is introduced into the recipient cell. The ssDNA can serve as the substrate for host recombination machinery and integrate into the host chromosome or if the DNA acquired is a plasmid, it has to establish itself to replicate independently (for review see (16–18)).

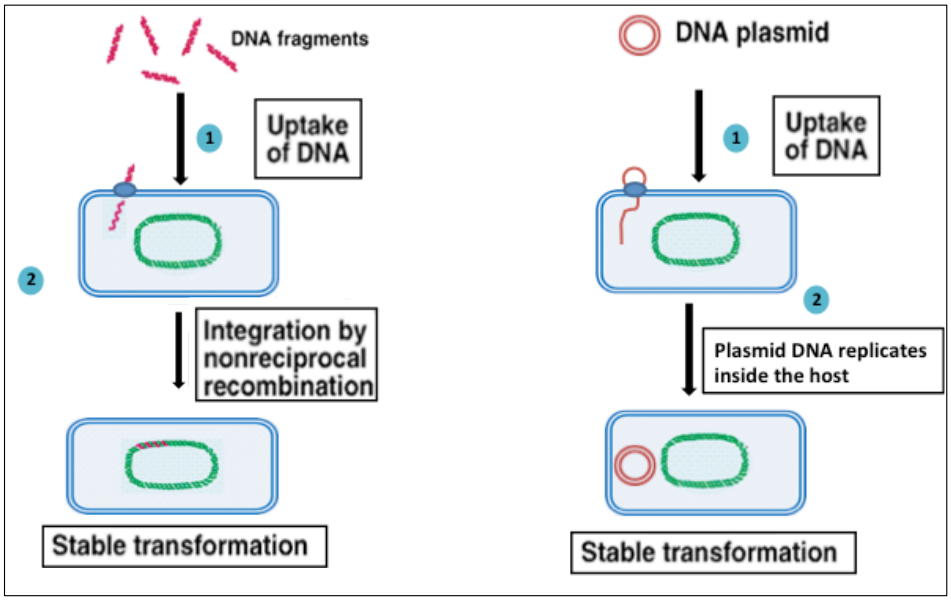


Figure I.1 **Schematic overview of the transformation process.** The host can take up DNA fragments or plasmid, independently replication extra-chromosomal elements. DNA fragments from other sources may either be recombinated or degraded as shown in step 2. Plasmids when uptaken by the host cells if are transformed, will replicate inside the host cell.

I.1.2 Transduction

Another mechanism causing LGT is transduction. A bacteriophage can be are responsible for transduction when it incorporate bacterial genomic DNA instead of, or addition to viral DNA in its capsid and subsequently infects a bacterial cell. Phages are extremely common in the environment and quite stable owing to their protective coat (for review see (17, 19)).

The first step of infection consists in the attachment of the phage with cognate receptor molecules on the surface of the bacterial cell. Phages attach to their host receptor via specific moieties known as anti-receptors. The specificity of phage anti-receptors is often flexible, allowing them to recognize several bacterial surface receptors. There are even examples known where phages contain multiple anti-receptors genes for different receptors (19, 20). Thus, flexibility in specificity may allow phages to invade host and some of them are able to mobilize bacterial genes and transfer them to other bacteria.

Transduction may be either generalized whereby any bacterial DNA can be packaged, an example of this is Coliphage P1, or it may be specialized. In the latter case, bacterial DNA that is packaged is situated near the site 0f prophage integration and one of the most studied example is Coliphage λ (17, 19, 21, 22).

The steps involved in transduction are schematically represented in **Figure I.2**. The process is as follows. The phage can enter the lysogenic cycle or the lytic cycle. The phage binds to the surface of a host bacterial cell and injects its genome into the cytoplasm. New phage genome and packaging proteins (i.e capsid, tail and tail fibers) are synthesized and new virions are assembled. Phage DNA and in some cases host DNA is incorporated into the phage heads .The infected cell is lysed and the transducing phage can injects its DNA into a new host. Injected DNA is incorporated into the chromosome of the infected cell by recombination.

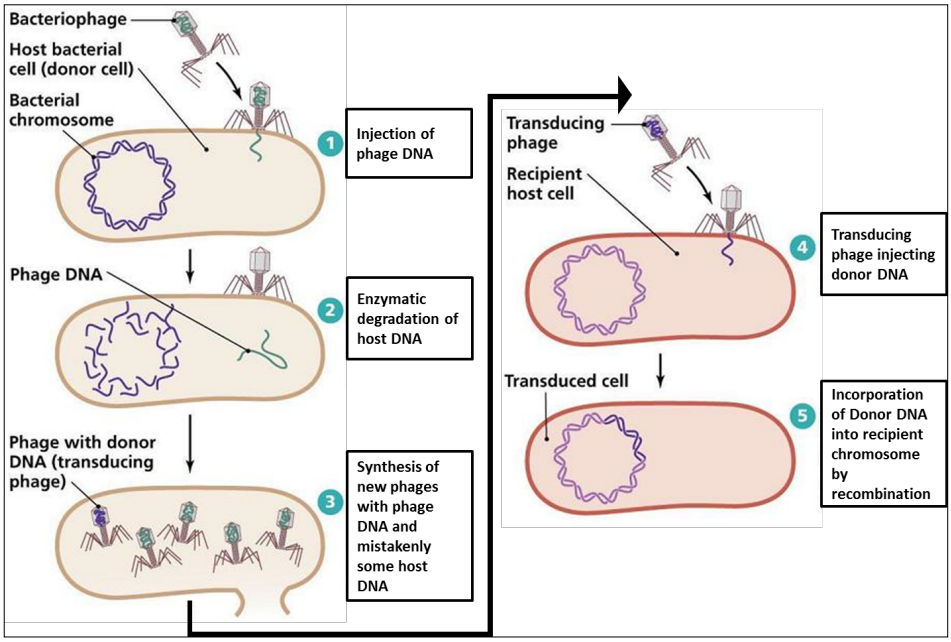


Figure I.2: **Schematic overview of the transduction process.** The first step of the infection of the bacterial cell by the phage, followed by the injection of the phage DNA. After the lytic cell ultimately, the transducing phage infects the recipient host cell, the donor DNA is integrated into the chromosome by recombination (see text for detail) .

I.1.3 Conjugation

Conjugation is the transfer of DNA from a donor to a recipient cell requiring contact between bacterial cells by a specialized multiprotein complex. Conjugation is a cornerstone of bacterial genetics and is also termed “bacterial sex”. Conjugation has the broadest range of host-transfer and it is also the most promiscuous mechanism of transfer of genetic material in comparison to transformation and transduction. Conjugation is widely seen in both Gram-negative (G-) and Gram-positive (G+) bacteria and is mediated by self-transmissible plasmids, conjugative transposons and integrative conjugative elements. (for review see (1, 3, 23))

Various aspects of the conjugation process is described later **section I.3**, **Figure I.3** gives a schematic overview of the steps involved in conjugation, which are as follows:

i) The donor cell initiates contact with the recipient cell through the mating pair formation (Mpf) system. Simultaneously, a protein called relaxase (with or without auxiliary proteins) cleaves the DNA of the conjugative elements in a site- and strand-specific manner within the origin of transfer (*ori*T) initiating thereby the

generation of ssDNA. The relaxase remains covalently attached to the 5'-DNA end, resulting in a complex known as the relaxosome. (see I.3.3.1)

ii) The 'coupling protein' (T4CP) links the relaxosome to the secretion pore formed by the Mpf system, through which the ssDNA passes into to the recipient cell. (see I.3.3.2)

iii) The ssDNA is converted into double-stranded DNA (dsDNA) in the recipient cell.

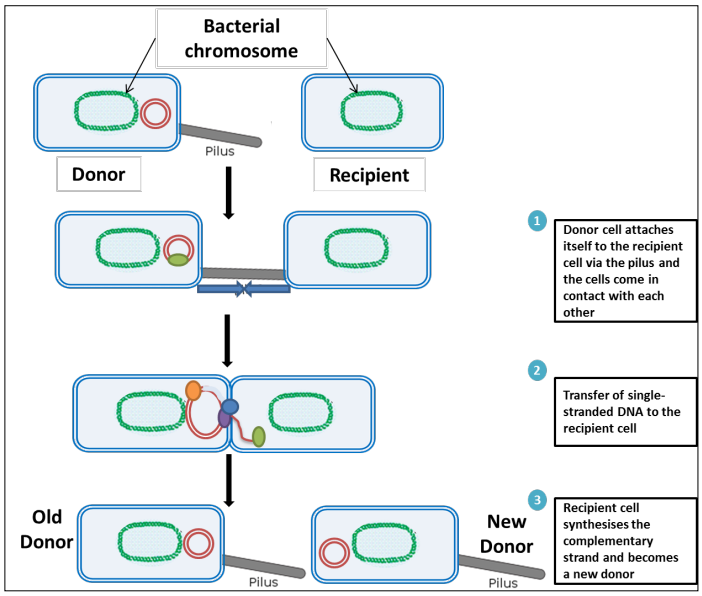


Figure I.3: A **schematic representation of conjugation**. . Conjugation is unidirectional transfer of DNA from the donor to the recipient cell requiring contact. Finally, the recipient cell becomes a new donor cell with its own copy of the plasmid.

I.2 Mobile Genetic Elements (MGEs)

Mobile genetic elements (MGEs) are segments of DNA that encode enzymes and other proteins that can mediate the movement of DNA within a genome (intracellular mobility) or between bacterial cells (intercellular mobility). There are several types of MGEs, many of these elements have been mentioned in the foregoing section. Below, we give a brief description for each type of MGE.

A) **Plasmids**—DNA elements that autonomously replicate apart from the host chromosome. Plasmids will be explained in greater detail in **section I.3**.

B) **Bacteriophages**—viruses that infect bacterial cells using a protein package, named capsid that contains genetic information. Usually the genetic information is DNA

though some RNA phages do exist. At low frequency, bacteriophages can accidently package segments of host DNA in their capsid and can inject this DNA into a new host, where it can recombine with the cellular chromosome (24).

C) **Transposons**—sections of DNA located between repeated sequences that can be excised and moved to a separate location via a site-specific recombination event involving the recombination enzyme called transposase. Intracellular movement of DNA is a property of promiscuously recombining transposons, which randomly recombine or 'jump' between replicons. Transposons can also 'hop' into phages and plasmids (3, 7, 25).

D) **Integrative and Conjugative Elements (ICEs)** are a diverse group of complex MGEs that encode the functions required to integrate into the bacterial host chromosome and to transfer themselves between cells via conjugation. ICEs include conjugative transposons and conjugative transposon-like genomic islands, as well as numerous unclassified MGEs. Genomic sequence analyses suggest that ICEs are widespread in bacteria and they probably contribute substantially to LGT. ICEs need to excise from a donor's chromosome into a circular form prior to transfer. Integration and excision of ICEs are recombination events catalyzed by serine or tyrosine integrases (Int) between short homologous sequences called attachment sites (*att*), on the circular element (*attP*) and the chromosome (*attB*) or flanking the integrated element (*attL* and *attR*), respectively (for review see (1, 3, 26)).

I.3 Plasmids

Plasmids are autonomous and independently replicating extra-chromosomal DNA elements. Although most plasmids are covalently circular DNA elements, some plasmids are linear. Joshua Lederberg coined the term 'plasmid' in 1952 to describe the process of bacterial conjugation, long before any plasmid structure had been seen (27). Plasmids can be of varying sizes from 2 to 300 kb. Plasmids are versatile vectors of biotechnological and industrial significance, which provide a means to transfer unrelated genetic material to a desired recipient cell. Plasmids may also serve as vehicles to transfer transposons and integrons. Thus, through plasmid conjugation

bacteria are exposed to a wide array of genes from the mobile gene pool. In some cases plasmids can contribute substantially to the total genetic content of a cell, representing more than 25% of the genetic content of the cell (7, 25, 28).

The general anatomy of a plasmid includes the essential 'backbone' of genes that have replicative and maintenance functions. In addition, plasmids may or may not code genes for transfer. Plasmids must replicate, control their copy number, and ensure their inheritance at each cell division by a process known as partitioning (for review see (29)). Most plasmids replicate by either the theta-type or rolling circle type of replication (for review see (30, 31)).

In theta-type replication, a replication initiator (Rep) protein recognizes the origin of the plasmid, binds to it, and facilitates the melting of the two strands. This step is followed by the recruitment of the host factors to the origin and commencement and elongation of DNA synthesis until the process terminates. Theta-type replication can be both unidirectional and bi-directional in nature. The term "theta" was used to describe the process of replication, as the DNA molecular were shaped like the Greek character "theta" Θ when visualized by electron microscopy (for review see (30)).

Other plasmids replicate via the rolling circle type of replication (RCR). These plasmids contain two replication origins, a double- stranded origin (*dso*) and a single-stranded origin (*sso*). The replication initiator (Rep) protein of RCR plasmids belongs to the relaxase family of proteins. The RCR Rep plasmid recognizes the *dso*, introduces a single nick in it, and becomes covalently attached to this strand. Next, replication is initiated at the exposed 3'-OH DNA group at the nick-site via other factors including DNA polymerase III, helicase, and ssDNA-binding proteins. Finally, termination of the replication cycle results in a newly copied, double-stranded molecule and a displaced, single-stranded molecule. The conversion of the ssDNA to a dsDNA molecule is mediated at the *sso* of the plasmid by the host factors, which results in the formation of complete double stranded molecule (for review see (31)). Thus, the replication

mechanism of RCR plasmids is mechanistically similar to the DNA processing that occurs during conjugation (see also **section I.3.3.1**).

Plasmids are workhorses whose mobility has opened up possibilities of genetics manipulation and also the root cause of the spread of antibiotic resistance genes or virulence genes among pathogenic bacteria. Based on their mobility, plasmids have been divided into three groups: non-mobilizable, mobilizable and conjugative (17, 24, 32).

Classification of plasmids based on their mobility

I.3.1 Non-mobilizable plasmid

Non-mobilizable plasmids, as the name indicates, are plasmids that cannot be transferred laterally to other cells via the process of conjugation or mobilization. However, other processes of LGT, like transformation and transduction, might mediate the transfer of these plasmids to other cells (32). Currently, there are many non-mobilizable plasmids known. Examples of non-mobilizable plasmids are *Bacillus subtilis* plasmid pTA1040 (33), *Burkholderia phymatum* STM815 plasmid pBPHY01 and *Ralstonia solanacearum* GMI1000 plasmid pGMI1000MP (32).

I.3.2 Mobilizable plasmids

Plasmids are called mobilizable, if they contain a minimal set of genes that allows them to be transmitted in the presence of a helper conjugative plasmid. The minimal set of genes, encode proteins that are important for the processing of the DNA, namely the relaxase and origin of transfer region on which the relaxase can act on (34). Examples of mobilizable plasmids are pTA1015 and pTA1060 isolated from *B. subtilis* (*natto*) strains (35, 36) or pUB110 found in *Staphylococcus aureus* which can be mobilized by pLS20 (37) or a *Streptococcal* plasmid pMV158 is mobilized by IncW R388 plasmid and IncP RP4 plasmid (38).

Moblizable plasmids are able to be conjugatively transferred, but do not encode all the functions necessary for self-transmission. Mobilizable plasmids seem to have a large impact in nature, including the spread of antibiotic resistance. Since plasmid mobilization is an almost universal procedure for gene spread among G+ and G- bacteria, a classification of plasmids according to their mobilization properties could be universal. Thus, based on their relaxases and the phylogenetic relationships among them, mobilizable plasmids have been divided into 4 different types, namely RSF1010, ColE1, pMV158 and CloDF13 families (32, 34).

In the next section, the process and mechanistic aspects of conjugation and mobilization are explained, as certain steps overlap.

I.3.3 Conjugative plasmid

Conjugative plasmids encode all the genes necessary to mediate their transfer from a donor to a recipient cell widely seen in plasmids found in both G+ and G- bacteria (for review see (1, 3, 23, 39–42)). The primary requisite of conjugation is the formation of intimate contact between the donor and the recipient cell. For G- bacteria the contact between the donor and recipient cell is arbitrated by mating pair formation (Mpf) proteins- the secretion channel as well as the pilus or other surface filaments (see review (40)). However, for G+ bacteria, no cell-cell contact has yet been identified, but may be mediated by surface adhesins.

I.3.3.1 DNA processing; formation of the relaxosome

The processing of substrate DNA for conjugative transfer is a widely conserved reaction among G- bacteria and unicellular G+ bacteria.

The origin of transfer or the *ori*T is the only *cis*-sequence required for transfer. The *ori*T along with the relaxase and other auxiliary proteins forms the transfer gene complex called the relaxosome. *ori*Ts usually are about 500 bps and contain the *nic* site. The *nic* site is the recognition site of about 10 nucleotides to which the relaxase binds and exerts the cleavage of the phosphodiester bond of a specific dinucleotide and attaches

itself to 5'-end of the nicked DNA strand. The transfer of DNA proceeds in a 5′→3′ direction. The region 5′ of *nic* site is referred to as upstream, or trailing region of a transferring stand, since it is the last portion of ssDNA strand to enter the recipient cell. The region 3′ of *nic* is referred to as downstream or the leading region. *ori*T regions are characterized by the following features. First, they are generally rich in A/T sequences facilitating negative super coiling of the DNA region. Second, they often contain direct and indirect repeats as well as intrinsic bends, which favor the binding of auxiliary proteins thereby altering the *ori*T structure locally (41).

Relaxase is the protein that binds to the *ori*T,and catalyzes the nicking reaction. The relaxase is a transesterase that preserves energy from cleavage of the phosphodiester bond of the transfer strand (T-strand). The relaxase with other auxiliary proteins form the relaxosome, thus, mediating the nicking reaction. The N-terminal domain of relaxases catalyzes the nicking reaction, while the C-terminal domain can have different activities; it can function as a DNA helicase, DNA primase or other functions (41, 43). Most but not all, relaxases contain in their N-terminal domain a conspicuous signature (called the 3H-motif), which can form a histidine triad.

Six different MOB families exist. Of these families, members of the MOB_P, MOB_Q, and MOB_V family contain only one tyrosine in their active sites. Members belonging to the MOB_F contain two Tyrosines in its active site. Finally, families MOB_H and MOB_C catalyze the reaction by an unknown mechanism (43).

I.3.3.2 Transfer of the DNA via the transferosome

Type IV secretion systems (T4SS) translocate DNA and proteins across the cell envelope (see reviews (44–47)). One type of T4SS is the conjugation system. As described above, relaxases with one or more auxiliary proteins form a relaxosome complex at *ori*T. The nucleoprotein complex is recruited to the cognate conjugation apparatus known as the transferosome, by interaction with a highly conserved ATPase termed as the substrate receptor or type IV coupling protein (T4CP). T4CP physically

interacts with the translocation channel, which is comprised of the mating pair formation (Mpf) proteins and then the DNA attached with the cognate relaxase is pumped into the recipient cell.

Coupling proteins (CPs)

Coupling proteins recruit the nucleoprotein complex to the Mpf assembly and are not involved in the DNA processing reactions or in pilin formation. So functionally, CPs constitute of usually two amino-terminal proximal regions, along with two transmembrane helices and a small periplasmic motif (in case of G- T4CP), and a large carboxy-terminal that is located in the cytoplasm. CPs include TraG (RP4), TrwB (R388) and VirD4 of T-DNA transfer (48).

Structural and energetic requirements for substrate translocation through the periplasm and outer membrane

Little is known about the architecture or mechanisms of action of type IV secretion channels or mpf channels in G+ bacteria. One of the most studied mating pair formation channel/system is the VirB1–VirB11 of the *Agrobacterium tumefaciens* VirB/D4 T4SS system with its interaction with the T-DNA. The transenvelope protein complex and the pili function in a co-ordinated manner, as a single hetero-protein complex, sanctioning the translocation (45, 46, 49).

Transenvelope Mpf structure

T4CP works in co-ordination with the Mpf complex, CP interacts with the latter, through VirB10 in case of T-DNA of *A. tumefaciens*. Through VirB10 CP harmonizes the passage of the T-strand, through the Mpf channel. Based on their functionality VirB proteins have been classified as putative channel subunits, which are divided into types inner-membrane (IM) proteins VirB6, VirB8 and VirB10, or outer-membrane (OM) proteins VirB3, VirB7 and VirB9, secondly ATPases, VirB4 and VirB11 providing energy to the aid the substrate transfer and finally, the pilin subunits VirB2 assembles

at the T-pilus in coalition with VirB5 and the VirB7 lipoprotein (for review see, (44, 45, 50, 51)).

Recapitulation

The formation of the relaxosome is an essential step of conjugation. Identification and characterization of the relaxosome components is required, in order to develop vectors based on conjugative plasmids for the modification of commercially, scientifically or clinically important strains.

I.4 Regulation of conjugation machinery

I.4.1 Regulation of conjugation in well-studied plasmid found in Gram-negative bacteria

The transfer process during conjugation can be divided into different stages: the mating pair formation, the relaxosome, and finally the transfer to the recipient cell. Inappropriate formation of such elaborated machineries can pose a burden on the host cell (for review see, (23, 52)). Conjugative elements on plasmids have evolved circuits to minimize the metabolic and phenotypic load on the host, while optimizing the benefits of the plasmid possessing of a transfer apparatus. Below, the basic principles of transcriptionally regulating circuits of some prototype conjugative elements are described.

I.4.1.1 The F-like transfer system (F-plasmid)

F-like plasmids all replicate in G- bacteria. The prototype of this family is the plasmid the thoroughly studied F-plasmid (100 kb in size), isolated from *E. coli* (39). Other plasmids belonging to this family are R1, a 94.7 kb plasmid isolated from *Salmonella paratyphi* (53) or plasmid R100 (89 kb) which was isolated from *Shigella flexneri* (54).

F-plasmid encodes for about 40 *tra* genes, forming the transfer region. Expression of the conjugation transfer genes of F-plasmid is controlled by the negative feedback

regulation mediated by the FinOP system and another loop controlled by three DNA-binding proteins TraM, TraJ and TraY (for review see, (23, 42, 52)) (see **Figure I.4**). The major control is the combination of FinO and FinP which together repress *traJ* mRNA. The fertility inhibition system (FinOP system) consists of *fin*P, which is an antisense RNA and the *fin*O product is a polypeptide. FinP binds to *traJ* mRNA preventing its translation, while FinO stabilises the FinP and *traJ* mRNA complex, promoting complex formation. TraJ, on the other hand, activates the promoter of conjugation P_{traY}, which encode the *tra* genes. TraY, creates its own regulatory circuit, by influencing both nicking and *tra* gene expression. TraY also activates P_{traM} (promoter of TraM), as they present as monocistronic operon.

Thus, the expression of *tra* genes is triggered when the balance between *traJ* mRNA and antisense FinP RNA transiently favours the sense RNA.

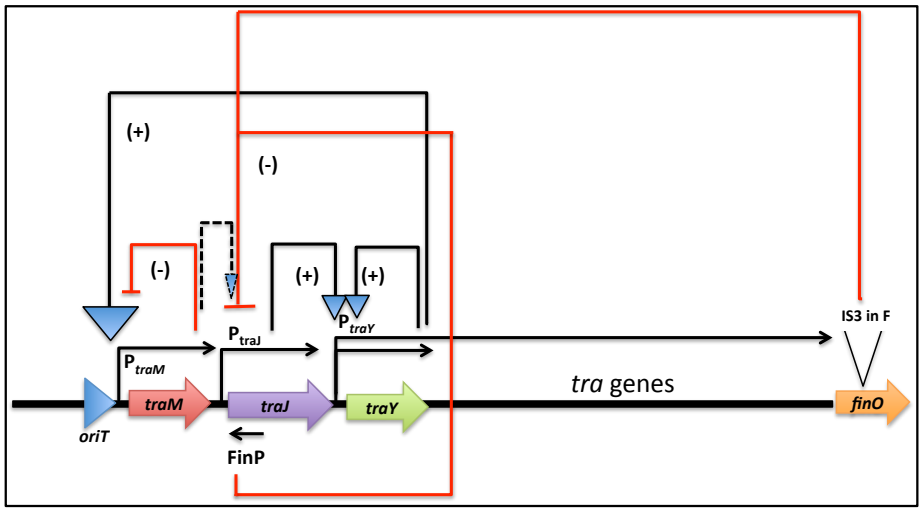


Fig. I.4 **A schematic overview of the control of F-transfer genes.** FinO and FinP together repress the translation of *traJ* mRNA, thus, negatively regulating system. While, TraJ activates P_Y (Promoter of *tra* genes) along with TraY itself. TraY also activates the promoter of *traM*. TraM inhibits its own production and influences P_{traJ}.

I.4.1.2 The IncP family (RK2)

IncP plasmids also are of G- origin. This family includes RK2 a 60 kb plasmid isolated from *Klebsiella aerogenes* (55), another 60 kb plasmid RP4 found in *Pseudomonas aeruginosa*) and a 54 kb plasmid called R64, isolated from *Salmonella typhimurium* (56); all isolated from G- bacteria. Expression of the IncP *tra* genes are controlled by

global regulators, as well as local auto-regulatory circuits, and co-ordination of transfer with other plasmid functions have evolved (for review see, (52, 57)).

The genes for conjugative apparatus of plasmid RK2 are organized in two blocks, designated as TraI encoding for *tra* genes and Tra2 coding for *trb* genes. The *tra* genes are involved in the formation of the relaxosome, while *trb* genes encode for proteins important for transfer.

Three global regulators control the system; the central control region encodes KorA and KorB, which is regulated by KorA. Further, KorA is required for the depression of *trb*A expression, providing a manner to shut down the transcription of *tra* and *trb* genes once the plasmid is established in the host. It also controls the expression of *trfA*, important for replication operon (58). KorB and TrbA repress the expression of *tra* genes directly (59). TrbA is a global repressor encoded in the Tra2 region (60). TrbA represses the transcription of both *trfA,* which encodes the replication protein and *trbB* is the protein of the Tra2 region as well as the expression of genes *tra*J, *tra*K and *tra*G (which form part of the Tra1 region) (58).

The assembly of the relaxosome is controlled by TraJ and TraK which bind at the *ori*T and repress the promoters of the region. TraJ binds specifically to *ori*T, recognizing a 10 bp palindrome in the right arm of the imperfect 19-bp inverted sequence repetition that is positioned upstream of *nic* (61). This nucleoprotein complex is recognized by TraI, the relaxase that cleaves at the *nic* site (62). While, TraK-*ori*T complex is important for the local unwinding of the DNA, although TraK is not essential for DNA processing (63).

Thus, in summary, KorA, KorB and TrbA function to globally regulate the system and locally the relaxosome formation, by its circuit involving TraJ and TraK.

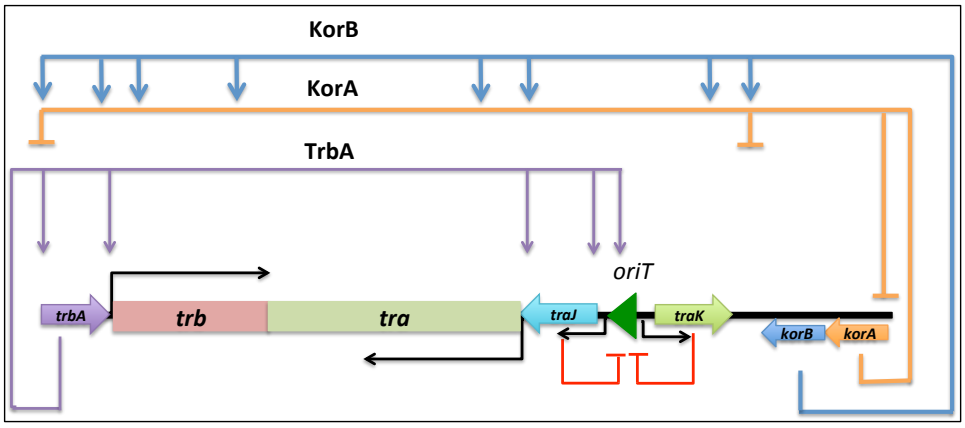


Fig. I.6 **A schematic representation of the control of IncP RK2-transfer genes.** Three are three main-regulators are KorA, KorB and TrbA. KorB and TrbA repress the expression of *tra* genes. KorA is important for shutting off the transcription at *tra* and *trb*, when the plasmid is established in the new host.

I.4.2 Plasmids of Gram-positive bacteria

Enterococcal plasmids like pAD1, pCF10 and pPD1 have been thoroughly studied. Expression of conjugation genes is activated after recipient-produced sex pheromones are absorbed by donor cells (for review see, (64–67)).

I.4.2.1 The pheromone-responding plasmid -pCF10

The regulatory control of the pCF10 conjugation genes is the combined action of a number of positive and negative regulators, a complex anti-termination system and a repressor of pheromone production activity (for review see, (66, 67)).

The genes involved in the regulation of pheromone induced conjugation are clustered in a 7 kb region present on pCF10, consisting of *prg*-pheromone response genes *prg*N-O-P-W-Z-Y-X-Q-R-S-T-A-B-C (68). Pheromones are short peptides composed of hydrophobic amino acids and cCF10 is the heptapeptide pheromone produced by both donor and recipient cells (69). cCF10 is encoded by a chromosomal gene. While, only the donor cells produce another signaling peptide iCF10, encoded on the plasmid pCF10 (a product of *prg*Q), which neutralizes the activity of cF10 (70). Plasmid pCF10 prevents the host from mating with plasmid-harboring cells by blocking the pheromone activity through the inhibition of its synthesis. In the un-induced state,

PrgY (*prg*Y is in regulation region of the plasmid) may be involved in the degrading or sequestering cCF10, inhibiting the pheromone activity (71). PrgX (encoded by *prg*X) along with iCF10 binds to the main conjugation promoter P_Q preventing the expression of conjugation genes. However, P_Q is not completely repressed it produces small transcript Q_S. Further, Q_S transcripts leads to the production of iCF10 peptide (70).

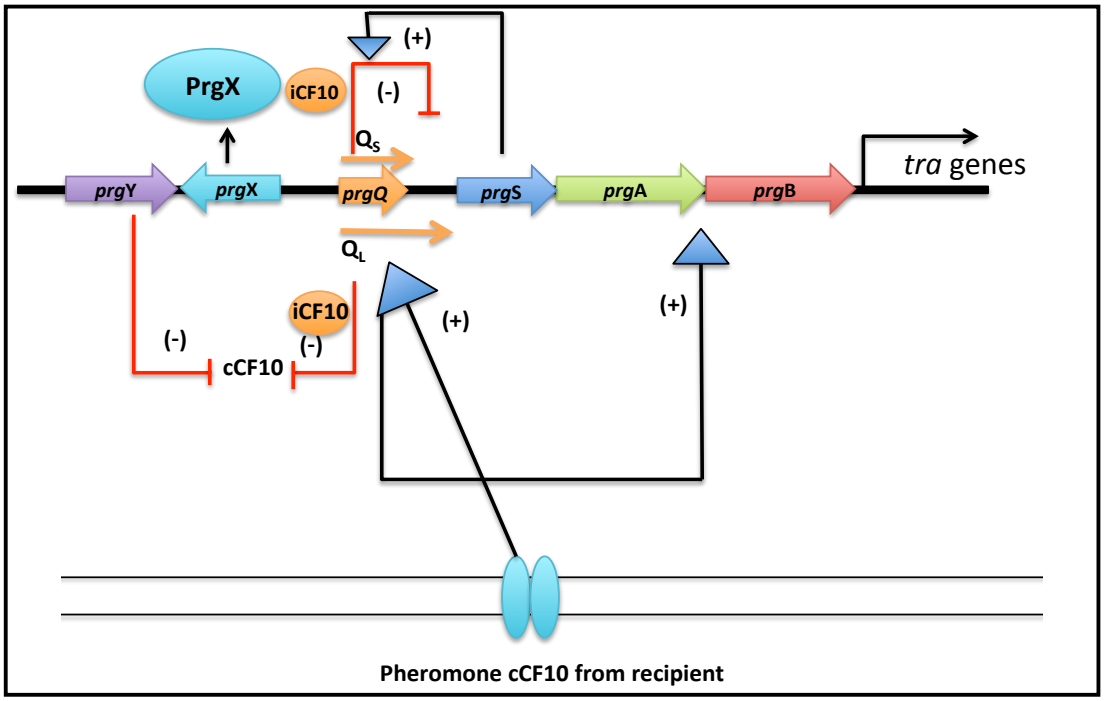


Fig. I.6 **A schematic representation of the control of pheromone regulated pCF10-transfer genes.** The negative regulation maintained by PrgX and iCF10. The uptake of the pheromone-produced cCF10 by the plasmid lacking recipient cells, which leads to the production Q_L instead of O_S, which extends into the *tra* genes (see text for details).

When induced by the production of cCF10 by plasmid-free cells, the donor (plasmid-harboring) cells express their *tra* genes. In addition, accumulation of pheromones indicates the close proximity of recipient cells (72). Donor cells respond to the presence of pheromone by the production of a probable membrane-spanning protein, Asc10 (encoded by *prg*B), which promotes aggregation with a recipient cell (68). The recipient pheromone cCF10 is internalized by oligopeptide binding proteins (OppA), encoded by *prg*Z in case of pCF10-harboring plasmids (73). When the pheromone enters a cell, the imported pheromone cCF10 interacts with PrgX complex initiated at P_Q, resulting in its de-repression. Activation of P_Q, by the pheromone-dependent manner, leads to the production of a longer product Q_L, instead of the Q_S product,

produced in the absence of pheromone. The most significant result of pheromone induction is to extend transcription from the *prg*Q promoter extending into the *tra* region and activating the production of the proteins for the plasmid transfer (74–77).

Thus, plasmid pCF10 regulatory circuit has evolved based on the interaction with the recipient cells, which helps to identify the plasmid-free from plasmid-harboring cells.

Conclusion

The process of conjugation and its transcriptional regulation have been well studied for plasmids found in G- bacteria, however, less is known from G+ bacteria, even though many G+ bacteria are important industrially and medically. It is important to understand the mechanism of the mobility of plasmids in G+ bacteria, as it would eventually permit the developed of specialized and better-adapted vectors for genetic engineering and the dispersion of genetically modified microorganisms for bioremediations and pest-control.

I.5 *Bacillus subtilis*

B. subtilis is a rod shaped, endospore-forming G+ bacteria, which was discovered in the 1800s. It was initially called ¨*Vibrio subtilis*¨ and was renamed as *Bacillus subtilis* by Ferdinand Cohn (78). *Bacillus subtilis* in particular has been intensely studied over many years and, as a consequence, is presently the best characterized G+ bacteria.

The advantages of using *B. subtilis* as the model organism to study conjugation are,

i) *B. subtilis* is referred to as a GRAS (generally regarded as safe) bacteria by US Food and Drug Administration.
ii) The whole genomic sequence of *B. subtilis* is available (79)
iii) *B. subtilis* can develop natural competence facilitating erroneously its genetic manipulation (80).

Additionally, the genus *Bacillus* includes both non-pathogenic and pathogenic species. Two pathogenic *Bacillus* species considered medically significant are: *B. anthracis,* which causes anthrax, and *B. cereus*, which is the causative agent of food poisoning. Other *Bacillus* species like *B. thuringenesis* produce insecticidal endotoxins that are used to control insect pests and hence have an agricultural importance. *B. amyloliquefaciens* is also source of natural antibiotic protein barnase, α-amylase, proteinase subtilisin, and *Bam*HI restriction enzyme. Thus, the genus *Bacillus* consists of many species having industrial, medical and agricultural implications (81, 82) making it a suitable model to study conjugation.

I.6 Plasmid pLS20

Plasmid pLS20 was first identified *Bacillus subtilis* IF03335. pLS20 is a large plasmid and its physical map with several type II restriction enzymes was constructed by Tanaka *et al* (83). Subsequently, Koehler and Thorne have shown that pLS20 is a conjugative plasmid and that it can mobilize several RCR-type plasmids (84). Itaya *et al* (85) marked pLS20 with a chloramphenicol gene and showed that pLS20 can conjugate in liquid medium.

Meijer *et al* (86) showed that replication origin of pLS20 is flanked by two divergently oriented genes *orf*A and *orf*B, which are not involved in replication. The minimal replicon contains several regions of dyad symmetry. Other features include, the presence of (i) DnaA boxes; (ii) an A/T-rich region containing several imperfect direct repeats; (iii) a replication terminator. Based on these features pLS20 was proposed to form a member of a new class of theta-replicating plasmids (86) .

pLS20 has an *alp*7A operon, constituting for two genes *alp*7A and *alp*7R (87). Alp7A is a type of bacterial actin and functionally assembles in the form of filaments and they manifest a dynamic stability (88). Alp7R negatively regulates the transcription of the *alp*7A operon. Alp7R dictates the activity and the cellular concentration of Alp7A, and it can interact with Alp7A (87). Thus, pLS20, which is present in the cell at low copy

number, and it, uses a dedicated mechanism involving the actin-like Alp7A protein for its segregation.

The sequencing and annotation of the plasmid pLS20cat was carried out in our laboratory. A schematic representation of the genetic map of plasmid pLS20 is shown in **Figure I.4**.

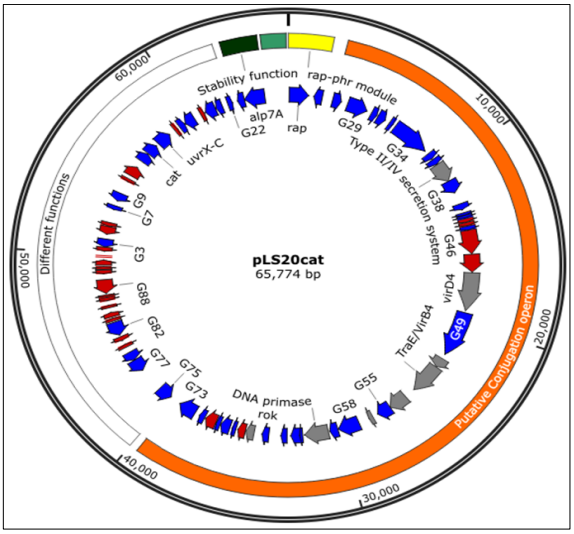


Figure I.7: **A schematic representation of the plasmid pLS20cat**, identifying the putative conjugative operon, marked in orange, along with ORFs with different functions labeled in white followed by the genes of stability, the *rap-phr* module noted in dark green and yellow respectively. The origin of replication is light green in colour.

In recent years, it has become clear that many differentiating process develop only in a sub-population of cells, even when the population of cells correspond to genetically identical cells and when they grow under identical conditions. This is referred to as bistable process. The development of competence is a bistable process. *B. subtilis* can develop competence, allowing the uptake of exogenous DNA, is regulated by various factors, repressor of *com*K (Rok) being one of them. Rok is a transcriptional repressor of competence genes, thus, limiting the size of the subpopulation, which develops competence. pLS20 possesses a *rok*-homologue, found to be shorter than the chromosomally encoded Rok protein. Ectopic expression of rok_{LS20}, leads to the inhibition of competence by binding to *com*K promoter. The analysis of the available databases, the several additional *rok*-like genes has been identified and thus, rok_{LS20} have been proposed to be prototype of a newly formed sub-group of nine *rok* genes(89).

Previously, we reported a global view of the regulatory circuitry of the pLS20 conjugation genes. A conjugation operon encompassing more than 40 genes is located next to a divergently oriented single gene, rco_{LS20}, which encodes the master regulator of conjugation responsible for keeping conjugation in the default "OFF" state. Activation of conjugation requires an anti-repressor, Rap_{LS20}, that belongs to the family of Rap proteins. Inactivation of the rap_{LS20} gene on pLS20 severely compromises conjugation, and conjugation was enhanced when rap_{LS20} was expressed from an ectopic locus. The activity of Rap_{LS20}, in turn, is regulated by a signaling peptide, Phr^*_{LS20}. The small phr_{LS20} gene, located immediately downstream of rap_{LS20}, encodes a pre-protein. After being secreted, Phr_{LS20} can be processed by a second proteolytic cleavage, resulting in generation of the functional pentapeptide, Phr^*_{LS20}, corresponding to the five C-terminal residues of Phr_{LS20}. When (re)imported, this peptide inactivates Rap_{LS20}. Therefore, activation of conjugation is ultimately regulated by the Phr^*_{LS20} signaling peptide. The Phr^*_{LS20} concentration will be relatively high or low when donor cells are predominantly surrounded by donor or recipient cells, respectively. Hence, conjugation will become activated particularly under conditions in which recipient cells are potentially present. In addition, Phr^*_{LS20} has a crucial role in returning conjugation to the default "OFF" state (90)

In the first part of thesis, we have looked in detail how conjugation is regulated by the master regulator Rco_{LS20}. Using, *in-vitro* and *in-vivo* analyses, it was found that the switching "OFF" system was regulated by three layers involves overlapping divergent promoters of different strengths regulating expression of the conjugation genes and the key transcriptional regulator Rco_{LS20}. The second layer involves a triple function of Rco_{LS20} being a repressor of the main conjugation promoter and an activator and repressor of its own promoter at low and high concentrations, respectively. The third level of regulation concerns formation of a DNA loop mediated by simultaneous binding of tetrameric Rco_{LS20} to two operators, one of which overlaps with the divergent promoters. The combination of these three layers of regulation in the same switch is complex, yet unique. It allows the main conjugation promoter to be tightly

repressed during conditions unfavorable to conjugation while maintaining the ability to accurately switch on the conjugation genes when appropriate conditions occur.

In the second part of the thesis, we genetically characterize the genes involved in the formation of the relaxosome, which is an essential step during conjugation and also localized the site of the origin of transfer of the plasmid pLS20 important for development of vectors based on conjugation. We found that proteins P56 and P57 are important along with the relaxase (encoded by gene 58), for the formation of the relaxosome.

2. OBJECTIVES

Our initial laboratory results identified the key players (Rco_{LS20}, Rap_{LS20} and Phr^*_{LS20}) involved in the regulation of conjugation-mediated by plasmid pLS20, providing a global view of the system. However, the molecular aspects of the regulatory system have to be understood in detail, further, one needs to identify the proteins involved in the formation of the relaxosome-an essential step of conjugation. Thus, the main objectives of this thesis have been listed below:

1. Understanding the genetic switch that regulates the expression of the conjugation genes.

2. Elucidating components of the relaxosome and followed by the identification of the origin of transfer *ori*T region of the plasmid pLS20.

3. MATERIALS AND METHODS

Bacterial strains, plasmids, media and oligonucleotides

Escherichia coli and *B. subtilis* strains were grown in Luria-Bertani (LB) liquid medium or on 1.5% LB agar plates (91). When appropriate, media were supplemented with the following antibiotics: ampicillin (100 µg/ml), erythromycin (1 and 150 µg/ml in *B. subtilis* and *E. coli*, respectively), chloramphenicol (5 µg/ml), spectinomycin (100 µg/ml), and kanamycin (10 µg/ml for *B. subtilis* and 30 µg/ml for *E. coli*). *B. subtilis* strains used were isogenic with *B. subtilis* strain 168 and are listed in **Table A.3**. Plasmids and oligonucleotides used are listed in **Table A.2** and **A.1**, respectively. All oligos were purchased from Isogen Life Science, The Netherlands.

Transformation

E. coli cells were transformed using standardized methods as described in Singh *et al* (92). For standard *B. subtilis* transformations, competent cells were prepared as described by Bron. Transformants were selected on LB agar plates with appropriate antibiotics.

Construction of plasmids and strains

DNA techniques were performed using standard molecular methods (91).The correctness of all constructs was verified by sequence analysis. The same strategy was used to construct *B.subtilis* strains containing a copy of *lacZ* fused to the entire or part of the rco_{LS20}-gene 28 intergenic DNA region. First, the region of DNA to be cloned was amplified using appropriate primers (**Table A.1**), purified, and digested with the appropriate restriction enzymes. Next, the fragment was used to prepare a ligation mixture together with the integration vector pDG1663 digested with the same enzymes. The ligation mixture was transformed into *E. coli* XL1-blue cells. The plasmid content of several ampicillin resistant transformants was checked and clones containing the insert with appropriate size and orientation were subjected to

DNA sequencing to verify the absence of mutations. The names of the pDG1663 derivatives and their characteristics are listed in **Table A.2**. Plasmid DNA of each pDG1663 derivative was used to transform competent *B. subtilis* 168 cell. Transformants were initially selected for resistance to erythromycin. Next, double cross-over events were distinguished from single cross-over events by selecting transformants sensitive to spectinomycin. The resulting *B. subtilis* strains containing a single copy of *lacZ* preceded by different regions of the rco_{LS20}-gene 28 region at the *thrC* locus of the *B. subtilis* chromosome are listed in **Table A.3**. Next, plasmid pLS20cat was introduced into the different *lacZ* fusion strains by conjugation. *B. subtilis* strain PKS9 contains a single copy of the rco_{LS20} gene under the control of the IPTG-inducible P_{spank} promoter at its *amyE* locus and this cassette is linked to the spectinomycin gene. Chromosomal DNA of strain PKS9 was used to transform competent cells of the various *lacZ* fusion strains in order to construct derivatives of the *lacZ* fusion 666 strains containing the P_{spank}-rco_{LS20} cassette.

The following strategy was used to construct a translational fusion of rco_{LS20} with his (89). The rco_{LS20} gene was amplified from pLS20cat by PCR using primers oPKS14N and oPKS8. The purified PCR product was digested with *Nco*I and *Sal*I and cloned into the vector pET28b+ digested with the same restriction enzymes to produce plasmid $pRco_{LS20}$-His. *B. subtilis* strain GR90 contains *the* rco_{LS20}-*his*(6) under the control of the Pspank promoter at the *amy*E locus. To construct this strain rco_{LS20}-*his*(6) was amplified from $pRco_{LS20}$-His by PCR using primers oGR3 and oGR4. The PCR product was digested with *NheI* and *SphI* and cloned into the vector pDR110 digested with the same enzymes to generate $pP_{spank}rco_{LS20}$-His. This plasmid was used to transform competent *B. subtilis* cells selecting for spectinomycin resistance. Double cross-over events were selected by loss of amylase gene.

Plasmid pLS20Δ*56-58* was constructed by the replacement of a kanamycin gene in place of genes *56, 57* and *58* in the plasmid *pLS20cat*. The upstream region of gene 56 was amplified using primers oGR56 and oGR57 and was named PCR UP 56 and

the downstream region of gene 58 was amplified using primers oGR58 and oGR59 and was named PCR DN 58. Plasmid pBEST501 was digested by *Bam*HI and *Sal*I to isolate the kanamycin gene and this purified kanamycin gene was ligated with PCR UP 56 digested with *Sal*I and PCR DN 58 digested with *Bam*HI. The ligation mixture was transformed into strain PKS 56, to give a strain GR148. The total DNA from the strain GR148 was checked for the insert and then the total DNA of strain GR148 was transformed into strain 168 to give strain GR149. The total DNA of GR149 was transformed accordingly to produce the strains GR150, GR197 and GR200, GR206. The transformed were checked by isolating the total DNA and carrying out PCR using the primer Ori_UP and Ori_Dn.

β-Galactosidase activity assays

Overnight cultures were diluted 100-fold into fresh medium and samples were taken at 45 min intervals for optical density reading (OD_{600}) and determining β-galactosidase activity as described previously (93).

Conjugation assays

Conjugation was carried out in liquid medium as described previously (90). The effect of ectopic expression of a given gene placed under the control of the inducible P_{spank} promoter on conjugation was studied as follows. Overnight cultures were diluted in prewarmed LB supplemented with IPTG at the indicated concentrations to an OD_{600} of ~0.05. Next, samples were taken at regular intervals to determine OD_{600} and were subjected to matings with proper recipient cells.

RNA isolation and RNA sequencing

Total RNA was isolated from exponentially growing cells by using RNeasy Mini Kit from Qiagen according to manufacturer´s protocol. RNA protect solution provided by

Qiagen was used to ensure the integrity of RNA during isolation and also to stop transcription at given time points. RNA sequencing and Bioinformatical analysis of RNAseq data was done as described previously (90).

Rco_{LS20}-His(6) purification

An overnight culture of *E. coli* BL21 (DE3) carrying plasmid pRco_{LS20}-His was used to inoculate (100-fold dilution) 1 litre of fresh LB medium containing 30 mg/ml kanamycin and incubated at 37°C with shaking. At an OD600 of 0.4, expression of *rco_{LS20}-his(6)* was induced by adding IPTG to a final concentration of 1 mM and growth was continued for 2 h. Cells were further processed as described previously (89). Purified protein (>95% pure) was dialysed against buffer B (20 mM Tris-HCl pH 8.0, 1 mM EDTA, 250 mM NaCl, 10 mM $MgCl_2$, 7mM β-mercaptoethanol, 50% v/v glycerol) and stored in aliquots at -80°C. Protein concentrations were determined by Bradford assay.

Gel retardation assays

Different fragments of intergenic regions between *gene 28* and *rco_{LS20}* were amplified by PCR using pLS20*cat* as template. The resulting PCR fragments were purified and equal concentrations (300 nM) were incubated on ice in binding buffer [20 mM Tris HCl pH 8, 1 711 mM EDTA, 5 mM MgCl2,0.5 mM DTT, 100 mM KCl, 10% (v/v) glycerol, 0.05 mg ml-1 BSA] without and with increasing amounts of purified Rco_{LS20}His(6) in a total volume of 16 µl. After careful mixing, samples were incubated for 20 min at 30°C, placed back on ice for 10 min, then loaded onto 2% agarose gel in 0.5XTBE. Electrophoresis was carried out in 0.5X TBE at 50 V at 4°C.Finally, the gel was stained with ethidium bromide, destained in 0.5XTBE and photographed with UV illumination (89).

Primer extension experiments

Determination of the transcription start sites by primer extension was performed essentially as described (94). In brief, total RNA (30 μg) was mixed with 4 pmol of end-labeled oligonucleotide that served as primer; the mixture was heated at 70°C for 5 min and allowed to anneal for 5 min at 23°C. The annealed RNA was ethanol precipitated, resuspended and primer extension was performed with 30 U of AMV reverse transcriptase (Promega) at 42°C,as recommended by the supplier. The extended cDNA products were analysed by electrophoresis on a denaturing 6% urea-polyacrylamide gel, in parallel with a DNA sequence ladder performed by chemical sequencing (95) of a DNA fragment encompassing the mapped promoters (see below). The primer used to map promoter P_c was 5'- ttctagttctttttacac, while that used for promoter P_r was 5'-tctctattgcccacttat. Oligonucleotides were end-labeled with [γ-^{32}P]-ATP and T4 polynucleotide kinase as recommended by the supplier (Biolabs). The 186 bp DNA fragment that served as sequence ladder was PCR amplified with primers 5'-acggtctagcgcttacaat and 5'-ttctagttctttttacac, the last one labeled at its 5' end.

DNase I Footprinting

DNaseI footprinting assay was carried out as described (96). The P_c/P_r promoter encompassing region was amplified by PCR using primers p28_Δ16 and Prom28UpBam, and pLS20cat as template. One of the ends was radio-labeled by digesting the fragment with *Bam*HI and subsequently filling in the end with exo-Klenow fragment in the presence of α- ^{32}P ATP.

Computer-assisted analysis

Presence of conserved motifs was searched by using motif-identification programs MEME (97) and BIOPROSPECTOR (98). Prediction of the static bending properties of DNA sequences was carried out by calculating the global 3D structure according to

the dinucleotide wedge model (99). All graphics work was done by using Adobe Photoshop CS2 and adobe illustrator. Graphs were plotted using Excel program.

Ultracentrifugation

Sedimentation velocity assay (SV)

Samples in 20 mM Tris-HCl, 250 mM NaCl, 10 mM MgCl2, 1 mM EDTA and 100 mM glycerol, pH 7.4, were loaded (320 µL) into analytical ultracentrifugation cells. The experiments were carried out at 43-48 krpm in a XL-I analytical ultracentrifuge (Beckman-Coulter Inc.) equipped with UV-VIS absorbance and Raleigh interference detection systems. Sedimentation profiles were recorded at 280 nm. Sedimentation coefficient distributions were calculated by least-squares boundary modelling of sedimentation velocity data using the continuous distribution c(s) Lamm equation model as implemented by SEDFIT 14.1 (100). Experimental *s* values were corrected to standard conditions (water, 20 °C, and infinite dilution) using the program SEDNTERP (101) to get the corresponding standard *s* values (*s*20, *w*).

Sedimentation equilibrium assay (SE)

Using the same experimental conditions as in the SV experiments, short columns (90 µL) SE experiments were carried out at speeds ranging from 7,000 to 10,000 rpm and at 280 nm. After the last equilibrium scan, a high-speed centrifugation run (48,000 rpm) was done to estimate the corresponding baseline offsets. Weight-average buoyant molecular weights of protein were determined by fitting a single species model to the experimental data using the Hetero Analysis program (102), and corrected for solvent composition and temperature with the program SEDNTERP (101).

4. RESULTS

Chapter I:

A complex genetic switch involving overlapping divergent promoters and DNA looping regulates expression of conjugation genes of the plasmid pLS20

As mentioned in the introduction, previous studies identified *rco*$_{LS20}$ (gene *27c*) as the gene encoding the protein that regulates the expression of the conjugation genes. These studies revealed that the default state of conjugation is the ¨OFF¨ state and that conjugation genes are activated by the antirepressor Rap$_{LS20}$ whose activity in turn is regulated by signaling peptide Phr*$_{LS20}$. Further, *rco*$_{LS20}$-gene *28* intergenic region contains the strong main conjugation promoter, P_c (see **Figure R.2** strain PKS3-fragment F_I$_C$-*lac*Z). Colonies of PKS3 (Fragment F_I$_C$ cloned in front of *lac*Z gene) were blue in presence of 5-bromo-4-chloro-indolyl β-D-galactopyranoside (Xgal) and quantitatively in the range of 300 and 500 Miller Units (MU). Introduction of pLS20 into strain PKS3, gave the strain PKS8, which is white in color after overnight incubation with Xgal containing LB plates. These results indicate that P_c is a rather strong promoter that does not seem to be regulated by host-encoded factors when grown under these conditions, but is regulated by an element present on the plasmid.

Despite these results, many questions remained unanswered.

- How Rco$_{LS20}$ regulates conjugation?
- Where the promoter of conjugation P_c and the promoter of repression P_r are?
- To what the operator sites does the repressor Rco$_{LS20}$ bind to?

R.1.1 pLS20*cat* conjugation efficiency and P_c promoter activity coincide

Under our laboratory conditions, efficient conjugation is limited to a narrow time window near the end of the exponential growth phase (90). If P_c is the main conjugation promoter it is expected that (i) its activity would generally be lower in the presence of pLS20*cat* and (ii) there would be a correlation between promoter P_c activity and the efficiency of conjugation. The following results show that this is indeed the case. In addition, when we used PKS8 as donor strain and simultaneously determined the kinetics of conjugation and promoter P_c activity we found that promoter P_c is only active during a rather short window of time near the end of the

exponential growth phase, which coincides with the period of high conjugation efficiency (**Fig. R.1**).

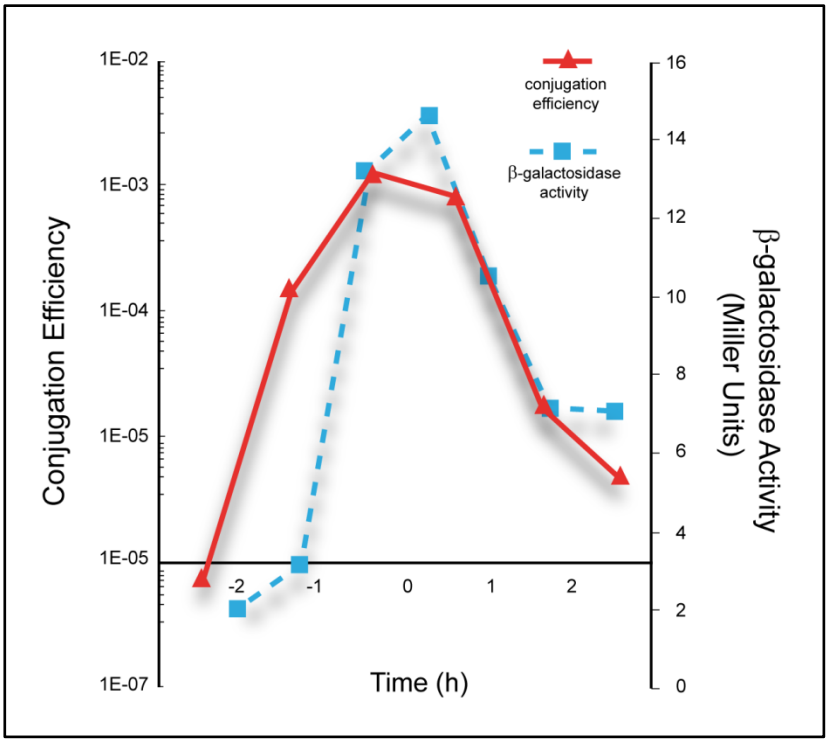


Figure R.1. **Correlation between the kinetics of P_c promoter activity and conjugation efficiencies of pLS20*cat* level** Overnight cultures of the strain **PKS8 (F_I$_c$-*lac*Z, pLS20*cat*)** and recipient strain PS110 were diluted to an OD_{600} of 0.05. Next, samples taken at different times were used to determine conjugation efficiency of pLS20c*at* by a standard conjugation protocol (continuous line), and promoter P_c activity by measuring β-gal activity (broken line). T=0 corresponds to the end of the exponential growth phase. The presented graph corresponds to a representative experiment. The experiment was carried out three times and the corresponding values differed by less than 10%.

R.1.2 Promoter P_c is located at an unusually large distance upstream of the first gene of the conjugation operon

Transcriptional fusions to several sub-fragments of the intergenic region were constructed and the entire intergenic region is referred to as Fragment I (or F_I). All

the fragments were cloned in the opposite orientation to analyze the divergent promoter of the *rco*$_{LS20}$ gene. For simplicity, the cloned fragments are indicated with oman letters. Fragments cloned in the orientation to analyze the conjugation and the *rco*$_{LS20}$ promoters are indicated with the extension "c" and "r", respectively.

A schematic representation of the different strains and fusions described in this work is given in **Figs. R.2B-C**.

These transcriptional fusions were carried out to delineate the position of the P_c promoter. As a first approach, we constructed strains containing *lacZ* fused to different subregions of Fragment I_c. Surprisingly, whereas no significant promoter activity was obtained with the strain having *lacZ* fused to Fragment II_c (strain GR10), the βG activities obtained with strains harboring *lacZ* fused to Fragment III_c, IV_c, Vc or VI_c were very similar to those obtained when *lacZ* was fused to Fragment I_c. These results show that promoter P_c is located at an unusually large distance of at least 350 bp upstream of gene *28*.

To delineate the position of the P_c promoter further we fused Fragment VIIc or VIIIc, which lack the 80 and 144 bp 5´-regions of Fragment I_c respectively, to *lacZ* (see **Fig. R.2.B**, strains GR68 and GR70). The reason why the 3´-endpoints of these constructs are downstream of the *Eco*RI site is explained below. Fragment $VIII_c$ (GR70) did not display promoter activity but the fusion based on Fragment VII_c (GR68) gave βG activities similar to that observed with Fragment I_c (**Fig. R.2.B**), showing that the 5´-located 63 bp region of Fragment VII_c contains (at least part of) the P_c promoter. This 63 bp region contains the sequence 5´- ttaaaaatttcactgaaatac-**TTtACA**-gttaaaaaaatgtc-TGt**TATctT**-3´, which constitutes a putative σ^A -dependent promoter endorsing several features characteristic for a strong promoter. First, the hexamer sequences 5´-TTtACA-3` and 5´-TATctT-3` are very similar to the consensus -35 (5´-TTGACA-3´) and -10 (5´-TATAAT-3´) sequences recognized by σ^A. Second, an optimal spacer length of 17 bp separates the putative -35 and -10 boxes. Third, the spacer contains the extended -10

motif (5´-TGn-3´, double underlined). Fourth, AT-rich tracts are located directly upstream of the predicted -35 box which are likely binding sites for the C-terminal domain of the RNA polymerase α-subunit. Additional evidence that this sequence constitutes the P_c promoter was obtained by primer extension analysis to determine the transcription start site. The detected extension product is shown in **Fig. R.4.B**. The position of the deduced transcription start site is located 6 bp downstream of the P_c core promoter sequences mentioned above (see **Fig. R.3.A**). The position of the transcription start site corroborates with RNAseq data, which provides a good estimation of the position of the transcriptional start site. Thus, total RNA isolated from pLS20cat-harboring cells were processed as described in Materials and Methods after which it was employed to generate cDNA libraries using a "directional RNA-seq" procedure that preserves information about the transcript's direction.

The schematic representation of the distribution and directionality of the reads presented in **Figure R.3.C** shows that the rightward-oriented transcripts, driving expression of gene *28* and downstream genes (shown in green), start close to the divergently oriented rco_{LS20} gene (shown in red). In fact, the most 5´ located reads detected by this method coincide with the transcription start site determined in the primer extension assay (not shown).

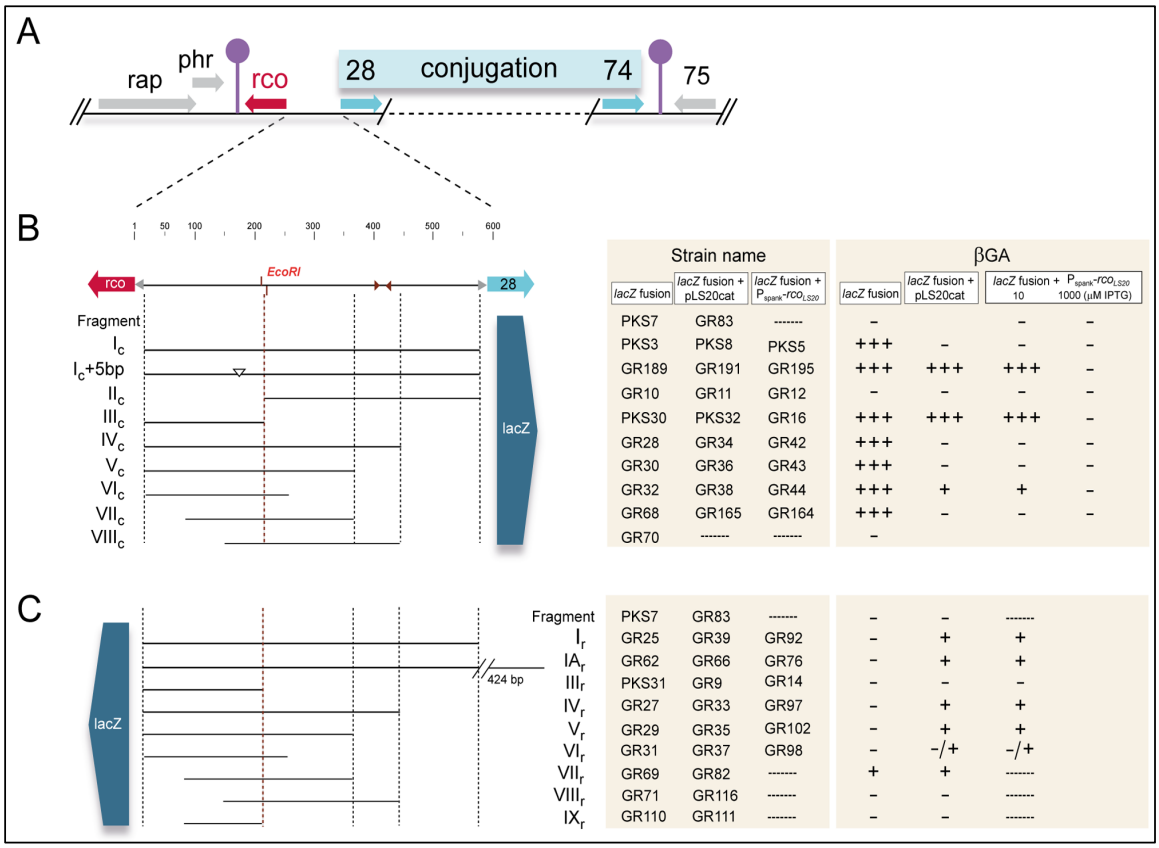


Figure R.2. **Genetic map of the studying the regulatory circuit of the plasmid pLS20*cat*.pLS20 conjugation region and overview of the transcriptional *lac*Z fusions used in these studies. A.** Map of the conjugation region of plasmid pLS20*cat*. Position and direction of genes and position of predicted transcriptional terminators are indicated with arrows and lollipop symbols, respectively. Panels **B** and **C** show a blow-up of the 600 bp rco_{LS20} -gene *28* intergenic region and the different fragments fused to *lac*Z. Fusions in (B) and (C) were used to study activity of promoters P_c and P_r, respectively. Features of the intergenic region are given on the top line. Numbers correspond to the bp position in this region. Names of the fragments cloned are indicated. Strains containing P_c-*lac*Z fusions in combination with the P_{spank}-rco_{LS20} cassette were grown on plates containing 10 μM or 1 mM IPTG. The symbols "+++", "+",and "-" reflect intense blue, pale blue, and white colonies after growth on X-gal containing plates. Color of colonies were observed after 16 and 48 hours of incubation at 37 ºC for strains containing pLS20cat or the P_{spank}-rco_{LS20} cassette, respectively.

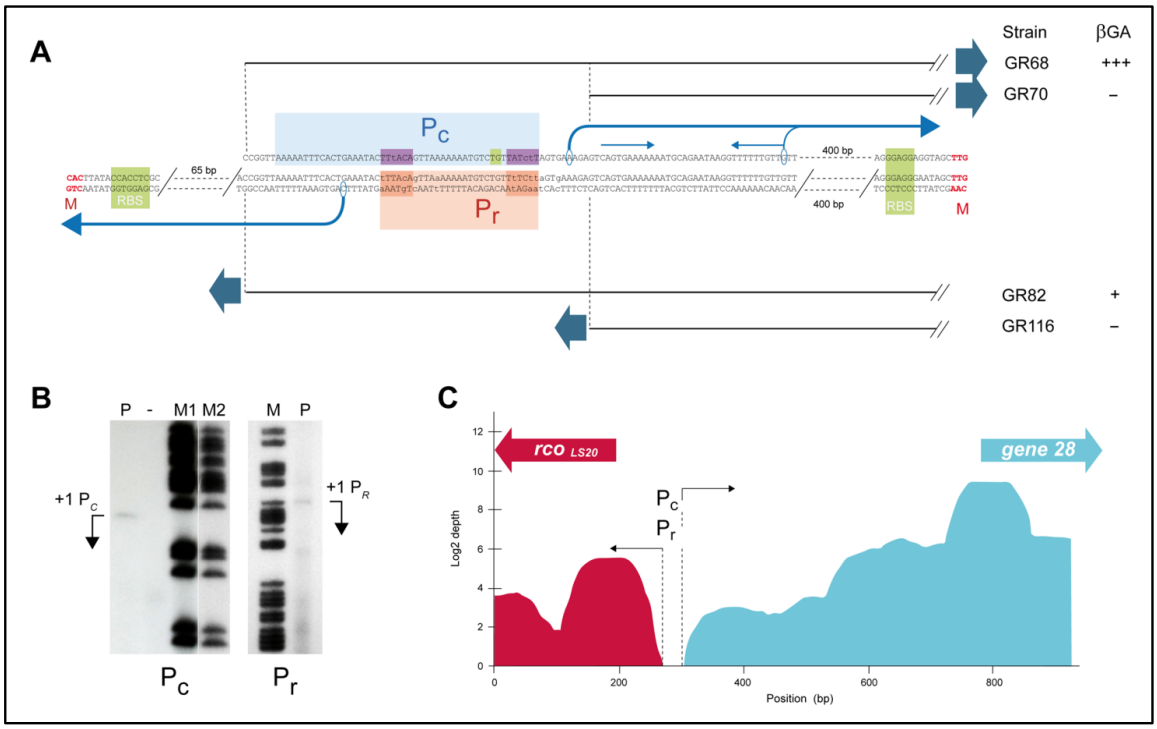


Figure R.3: **Promoter P_c is located 461 bp upstream of the start codon of gene 28 and overlaps with the divergently oriented P_r promoter. A**. Determination of promoter P_c and P_r sequences by deletion analysis and primer extension. pLS20*cat* containing cells harvested at the end of the exponential growth phase were processed to isolate their total RNA, which was used in primer extension assays as described in Materials and Methods. Promoter P_c. The lines above the sequence indicated the 5´ end points of the transcriptional lacZ fusions present in strains GR68 and GR70, displaying and not displaying promoter activity, respectively. The core promoter and putative upstream UP elements of are indicated by a light blue box; the -35 and -10 hexamers, and the extended -10 motif are indicated with dark blue and green boxes, respectively. The transcription start site determined by primer extension is shown with a black bent arrow. The thin grey bent arrow corresponds to the 3´ end point of the smaller extension product that corresponds to the start of an inverted repeat which is indicated with blue arrows above the sequence. Promoter P_r. The lines below the sequence indicate the 3´ end points of the transcriptional fusions with lacZ reporter present in strains GR82 and GR116, displaying and not displaying promoter activity, respectively. The deduced position of the P_r core promoter and the -35 and -10 boxes are indicated with orange and red boxes, respectively (see text). The transcription start site determined by primer extension is shown with a black bent arrow. **B.** Primer extension to determine the transcription start site of promoter P_c (left panel) and P_r (right panel). The cDNA products of the primer extension reactions are indicated with bent arrows (lane P). Free lane in which no sample was run is indicated with "-". Lanes M, M1 and M2 correspond to [G+A] chemical sequencing reactions of a short 230 bp DNA fragment corresponding to the studied pLS20*cat* region obtained by PCR amplification as described in Materials and Methods. In the case of the P_c promoter, a smaller extension product giving a stronger signal was observed 37 bp downstream of the extension product shown. Although producing a weaker signal, the longer extension product shown most likely reflects the correct transcription start site based on the following arguments. First, it is known that AMV reverse transcriptase prematurely terminates cDNA synthesis when reaching a stem loop in the RNA, and that the prematurely terminated molecules map at the bottom of the secondary structure (94, 95)The position of the strong signal coincides with the 3´ end of an inverted repeat (indicated in Fig. 3A). Second, no putative core promoter sequences are evident upstream of the 5´ position of the shorter extension product. Third, if the stronger signal corresponds to the transcription start site, the responsible promoter would be present on Fragment $VIII_c$ used for the transcriptional lacZ fusion in strain GR70. However, no promoter activity was observed with this strain (see text). And fourth, the transcription start site based on the longer extension product corroborates the RNAseq data. **C.** Schematic overview of RNAseq expression data of pLS20*cat* genes rco_{LS20} and *28* under conditions with (top panel) and without overexpression of rco_{LS20} (lower panel). The amount of right and leftward "reads", given in green and red, respectively, are presented on a log2 scale. The positions of the divergently oriented genes rco_{LS20} and *28* are indicated on the top with a red and green arrow, respectively. Black arrows indicate the approximate start of the divergent transcripts driven by the P_c and P_r promoters.

R.1.3 The *rco$_{LS20}$-28* intergenic region contains the weak P_r promoter that is activated and repressed at low or high Rco_{LS20} concentrations, respectively

As for P_c, we constructed *lacZ* fusion strains to characterize the divergently oriented P_r promoter responsible for expression of Rco_{LS20}. Surprisingly, no promoter activity was observed when *lacZ* was fused to the 570 bp Fragment I_r (strain GR25, **Fig. R.2.C**). One possibility could be that promoter P_r is located even further upstream. This does not seem to be the case however, because Fragment IA_r, corresponding to the 1,014 bp region upstream of *rco$_{LS20}$* (strain GR62), also did not provide detectable levels of promoter activity. After obtaining these negative results, we introduced pLS20cat into these strains to study whether it encodes a protein that might be required to activate promoter P_r. Colonies of the resulting pLS20cat-harboring strains GR39 (F_I_r-*lacZ*) and GR66 (F_I_{Ar}-*lacZ*) turned pale blue when grown on Xgal-containing plates (**Fig. R.2.C**), indicating that indeed pLS20cat provides a protein that activates the P_r promoter. In addition, the results show that the P_r promoter is located on Fragment I_r.

We then considered the possibility that Rco_{LS20} might be responsible for activating its own promoter. To test this, we engineered strain GR92 that contains the F_I_r-*lac*Z fusion combined with the cassette in which expression of *rco$_{LS20}$* is under the control of the IPTG inducible P_{spank} promoter. Colonies of strain GR92 were white when grown on agar plates containing only Xgal, but turned pale blue when the plates contained also low levels of IPTG. These results demonstrate that Rco_{LS20} activates its own promoter. In addition, the fact that colonies only developed a pale blue color suggests that promoter P_r is weaker than P_c. To test this more directly, we measured P_r promoter activities at late-exponential growth phase using strain GR92 grown at different levels of Rco_{LS20} induction (**Table T.2**). Interestingly, maximum P_r promoter activity was obtained when cells were grown in the presence of 50 µM IPTG. Promoter P_r activity decreased at higher IPTG concentrations and equaled background levels in the

presence of 1 mM of IPTG, indicating that Rco_{LS20} represses its own promoter at higher concentrations. Together, these results show that P_r is a weak promoter whose strength is several hundred folds weaker than that of P_c. The results also show that Rco_{LS20} has a triple function.

First, low levels of Rco_{LS20} are required to activate its own promoter P_r; second, at higher concentrations Rco_{LS20} represses its own promoter; and third, Rco_{LS20} is responsible for repression of the oppositely oriented P_c promoter. This triple function of Rco_{LS20} is likely to have important consequences for regulation of the conjugation genes (see **Discussion Chapter I**). It is worth mentioning that whereas maximum activation of the P_r promoter was achieved when *rco_{LS20}* was induced from the P_{spank} promoter at 50 mM IPTG, efficient repression of the P_c promoter was observed by inducing *rco_{LS20}* with as low as 10 mM IPTG. Finally, the results obtained show that Rco_{LS20} is the only pLS20cat protein required for activation and repression of the P_r and P_c promoters.

Table T.2: P_r promoter activity at different induction levels of Rco_{LS20}

IPTG (µM)	βGA (MU)
0	< 0.1
10	0.2
20	0.7
50	1.2
100	0.4
200	0.3
500	0.2
1.000	<0.1

Overnight culture of GR92 cells grown in the absence of IPTG was diluted 100-fold in fresh prewarmed LB medium containing the indicated amount of IPTG at 37 ºC. βGA was determined for samples withdrawn at an OD_{600}= ~ 1.0. Background levels obtained with negative control strain PKS7 were <0.1. βGA, β galactosidase activity

R.1.4 The divergent P_r and P_c promoters overlap

As a first approach to determine the position of the P_r promoter we constructed strains containing *lacZ* gene preceded by different subregions of Fragment I_r combined with pLS20cat to provide Rco_{LS20} in *trans*. The transcriptional regulator Rco_{LS20} is a DNA binding protein. Therefore, a lack of P_r promoter activity in the reporter assay can be due to the absence of (part of) the P_r promoter or the Rco_{LS20} 273 binding sites required for activation of P_r. Since activator proteins generally bind upstream of promoters, we tested constructs having deletions at the 3´ end of Fragment I_r (i.e. flanking the *rco_{LS20}* gene). Promoter P_r activity was detected when *lacZ* was fused to Fragment VII_r (strain GR82), but not when it was fused to Fragment $VIII_r$ (strain GR116) (**Figs. R.2.C and R.3.B)**. These results suggested that promoter P_r would be (partially) located on the 63 bp 5´ region of Fragment VII. Interestingly, the divergently oriented P_c promoter is located on this same 63 bp region (see above, **Fig. R.3.A**). In a complementary approach, we determined the transcriptional start site of promoter P_r by primer extension (**Fig. R.3.B**). The determined transcription start site of promoter P_r is positioned 6 bp upstream of the -35 box of the P_c promoter (see **Fig. R.3.A**). This implies that promoter P_r overlaps with the P_c promoter. P_r is a weak promoter whose activity requires Rco_{LS20}. It is therefore unlikely that the -35 and -10 boxes will be very similar to the consensus sequences. The following two sequences that may constitute a s^A -dependent promoter are located upstream of the determined P_r transcription start site: (i) [5´-aaGAtA- 17bp -TgTAAa-3`] and (ii) [5´-aTaACA-18 bp-aAgtAT-3`] (mismatches with respect to consensus -35 (5´-TTGACA-3´) and -10 boxes (5´-TATAAT-3´) given in lower case, see **Fig. R.3.A**). The position of the determined transcription start is optimally spaced with respect to the first but not the second possible promoter sequence. Therefore, we favor the first sequence to correspond to the P_r promoter. Interestingly, this would imply that the positions of the -10 and -35 boxes correspond exactly to the -35 and -10 boxes, respectively, of the divergently oriented P_c promoter. The results of the RNAseq experiments presented in **Figure R.3.C** supports the conclusion that the P_r and P_c promoters overlap. RNA transcripts mapped against the entire intergenic region except for a small region that is located near the start of the

rco$_{LS20}$ gene. The divergent promoters P_r and P_c, responsible for the left- (red) and rightward (green) oriented transcripts, respectively, must both be located in the small nontranscribed region which corresponds to the position of the P_r/P_c promoters according to their transcriptional start sites determined by primer extension.

In summary, results obtained by a combination of different approaches demonstrate that divergent P_c and P_r promoters overlap, if not coincide.

R.1.5 *In-vivo* evidence that Rco$_{LS20}$ binds to two operator sites; one of them, -located more than 85 bp downstream of P_c-, is required for efficient regulation of promoters P_c and P_r

Rco$_{LS20}$ belongs to the Xre-family of transcriptional regulators and is predicted to contain a Helix-Turn-Helix (HTH) DNA binding motif in its N-terminal region. It is therefore likely that Rco$_{LS20}$ will exert its transcriptional regulatory effects on P_r and P_c by binding to DNA sequences. We employed the following *in vivo* approach to gain insights into the location of the Rco$_{LS20}$ binding sites. Either pLS20cat or the P_{spank}-*rco*$_{LS20}$ cassette was introduced into the various *lacZ* fusion strains (see **Fig. R.2**). The resulting strains were then grown on Xgal containing LB plates, -supplemented with or without 10 µM of IPTG for strains containing the P_{spank}-*rco*$_{LS20}$ cassette-, and expression of the different *lacZ* fusions in response to Rco$_{LS20}$ was screened by the color of their colonies.

A schematic summary of the results obtained for promoter P_c is given in **Figure R.2.B**. In agreement with results presented above, the strain harboring *lacZ* fused to Fragment I_c (PKS3) displayed high P_c promoter activity, but no promoter activity was detected when Rco$_{LS20}$ was provided in *trans* (strains PKS5 and PKS8). Efficient Rco$_{LS20}$-mediated repression of the P_c promoter was lost however when *lacZ* was fused to Fragment III_c (strains GR16 and PKS32). This strongly indicates that an Rco$_{LS20}$ operator

site is located on the 368 bp Fragment II_c and that this operator, which would be located at least 85 bp downstream of the P_c promoter, is crucial for efficient repression of the P_c promoter. Fragment II_c contains an inverted repeated sequence (5´-ATCAAAATCAtgctgcaactTGGTTTTGAT-3´). To test whether this region constitutes an Rco_{LS20} operator site we constructed *lacZ* fusions to Fragments IV_c or V_c, and also engineered derivatives of these two strains containing pLS20cat or P_{spank}*rco*$_{LS20}$. The 5´ ends of these Fragments are located up- or downstream of the inverted repeat (see **Fig. R.2.B**). The P_c promoter present on Fragment IV_c and V_c was efficiently repressed by Rco_{LS20}, indicating that the Rco_{LS20} operator site is present on Fragment V and not on the 212 bp region immediately upstream of gene *28* containing the mentioned inverted repeat. Efficient Rco_{LS20}-mediated repression of promoter P_c was not observed for the *lacZ* fusion based on Fragment III_c (see above). Together these *in vivo* results strongly indicate that an ~ 160 bp region, located 85 bp downstream of Pc, contains an Rco_{LS20} operator site that is required for efficient repression of this promoter. We name this operator site O_I.

Results described above show that promoter P_c was not repressed by Rco_{LS20} when the *lacZ* fusion was based on Fragment III_c (strain GR16) and cells were grown in the presence of 10 µM IPTG. Interestingly though, promoter P_c in strain GR16 was efficiently repressed when the concentration of IPTG was increased to 1 mM (see **Fig. R.2.B**). This indicates that another Rco_{LS20} operator site is present on the 201 bp Fragment III_c. We name this operator site O_{II}.

Next, we used the same strategy to delineate the regions required for activation of the divergent P_r promoter. The results of these analyses are summarized in **Figure R.2.C.** Interestingly, the region required for efficient repression of P_c by Rco_{LS20}, is also required for Rco_{LS20}-mediated activation of promoter P_r. Thus, Rco_{LS20} activated the P_r promoter when *lacZ* was fused to Fragments IV_r or V_r (strains GR97/GR33 and GR102/GR35, respectively) but not when it was fused to Fragment III_r (strains GR14/GR9, **Fig. R.2.C**). In summary, the *in vivo* results obtained provide strong

evidence that one Rco_{LS20} operator, O_I, is located in an ~160 bp region located 85 bp downstream of promoter P_c, and that this operator is crucial for proper repression and activation of promoters P_c and P_r, respectively. In addition, the results indicate the presence of another Rco_{LS20} operator, O_{II}, which would be located near promoters P_c and P_r.

R.1.6 *In-vitro* approaches show that Rco_{LS20} binds cooperatively to multiple binding sites present in operators O_I and O_{II}

To study the position of the Rco_{LS20} binding sites in more detail we purified Rco_{LS20} and used it in Electrophoretic Mobility Shift Assays (EMSA). To facilitate purification, we constructed an *E. coli* strain that expresses an Rco_{LS20}-His(6) tagged fusion protein. The *his(6)*-tag was placed at the C-terminus because Rco_{LS20} contains a predicted Helix-Turn-Helix DNA binding motif close to its N-terminus. The following result demonstrates that the Rco_{LS20}-His(6) protein is functional *in vivo*. We constructed *B. subtilis* strain GR90 in which expression of *rco_{LS20}-his(6)* gene is placed under the control of the inducible P_{spank} promoter, and which also contains the F_I_c-*lacZ* reporter fusion. The activity of promoter P_c in strain GR90 was repressed in an IPTG-dependent manner similar to that observed for strain PKS5 containing an inducible copy of native *rco_{LS20}*.

The *in vivo* transcriptional fusion results presented above indicated the presence of two operators. One of them, operator O_{II}, located near promoters P_c/P_r, and another one, operator O_I, present in an ~160 bp region about 85 bp downstream of P_c. In addition, this analysis indicated that the ~200 bp region immediately upstream of gene *28* does not contain Rco_{LS20} binding sites. Accordingly, we began analyzing binding of Rco_{LS20} to Fragments X (200 bp region upstream gene 28), III (expected to contain operator O_{II}) and XII (expected to contain operator O_I) (see **Fig. R.4**). Control experiments showed that Rco_{LS20} did not bind to an unrelated 550 bp fragment of pLS20cat (not shown). Independent of the concentrations used, Rco_{LS20} did not bind to

Fragment X (**Fig. R.4.B**). Together with the *in vivo* data presented above, this provides strong evidence that this region does not contain Rco_{LS20} binding sites. Also in agreement with the *in vivo* data, Rco_{LS20} bound to both Fragment III and Fragment XII (**Fig. R.4.B**). Interestingly, the retardation patterns obtained for these fragments were similar, and resulted in the appearance of a maximum of two retarded species. The observation that the two retarded species were already present at low Rco_{LS20} concentrations when the majority of the DNA migrated to the position of unbound DNA, strongly indicates that Rco_{LS20} binds cooperatively to at least two binding sites present in each operator. In addition, the observation that DNA fragments entered the gel even at very high protein concentrations indicates that Rco_{LS20} binds to specific sites and that it does not spread along the DNA. To delineate the O_I and O_{II} regions further we used overlapping and subregions of Fragments III and XII as probes. Fragment IIIA (130 bp containing promoters P_c/P_r) and Fragment XIIA (125 bp) both produced up to two shifts, and Rco_{LS20} did not bind to the 46 bp region that separates these two fragments. This latter conclusion is based on comparison of gel retardations obtained with fragments XI and XII. We next analyzed binding of Rco_{LS20} to Fragments I, IV and V that encompass both operators. These fragments gave similar retardation patterns. Interestingly, in these cases, Rco_{LS20} binding resulted in the appearance of four retarded species. All four of these retarded species could be detected at low Rco_{LS20} concentrations when most of the fragment had not bound Rco_{LS20}, indicating that Rco_{LS20} binds cooperatively to multiple sites on these fragments.

To search for the presence of conserved motifs in the two Rco_{LS20} operators we used the motif-identification programs MEME (103) and BIOPROSPECTOR(98). These analyses revealed the identification of an 8 bp conserved motif that is present seven times in O_{II} (Fragment III_A), and four times in O_I (Fragment XII_A). We named the seven motifs identified in the O_{II} operator a1-a7, and the four motifs in the OI operator b1-b4 (see **Fig. R.5**). Whereas motifs b1 to b4 are all located on the lower strand, motifs a1-a7 are located on the upper strand, except motif a3. It is worth mentioning some characteristics of motifs a1 to a7. First, motif a5 overlaps with the P_c/P_r core

promoter sequences, and motifs a1-a4 and a6-a7 flank them. Second, motifs a1 and a7 form part of a 13 bp direct repeat. Third, motifs a1 and a3 form an inverted repeat. Fourth, the oppositely oriented motifs a3 and a4 overlap in a region that has an inverted repeat (5´TTTCAgTGAAA-3´). Evidence that the identified motif constitutes (part of) the binding site for Rco_{LS20} was obtained by DNase I footprinting and mutational analysis. Thus, gel retardation assays showed that binding of Rco_{LS20} is affected in probes containing alterations in one or two motifs in either operator. For instance, Rco_{LS20} did not bind to a derivative of Fragment III_A containing mutations in both motifs a1 and a7; and binding was affected when only motif a7 was mutated. Similarly, mutation of motif b1 or b4 resulted in the appearance of only one retarded species instead of two observed for corresponding fragments without mutations (**Fig. R.5.B**). In summary, the results obtained show that the intergenic *rco*$_{LS20}$-gene *28* region contains two Rco_{LS20} operators that are separated by 75 bp. Operator O_{II} overlaps with promoters P_r/P_c and the other region is located 75 bp towards the direction of gene *28*.

Each region contains repeats of a motif whose consensus sequence is 5´-CAGTGAAA-3´ and which forms (part of) the binding site of Rco_{LS20}. Motifs in O_I are located on the lower strand, and except for one, motifs in O_{II} are located on the upper strand.

Binding of Rco_{LS20} to operators O_I and O_{II} was confirmed by DNase I footprinting. The results presented in **Figure R.6** confirm that Rco_{LS20} binds to a region that overlaps with the Pr/P_c promoters and to another region located about 75 bp downstream of the P_c promoter.

The combined *in vitro* results are in line with the *in vivo* results presented above.

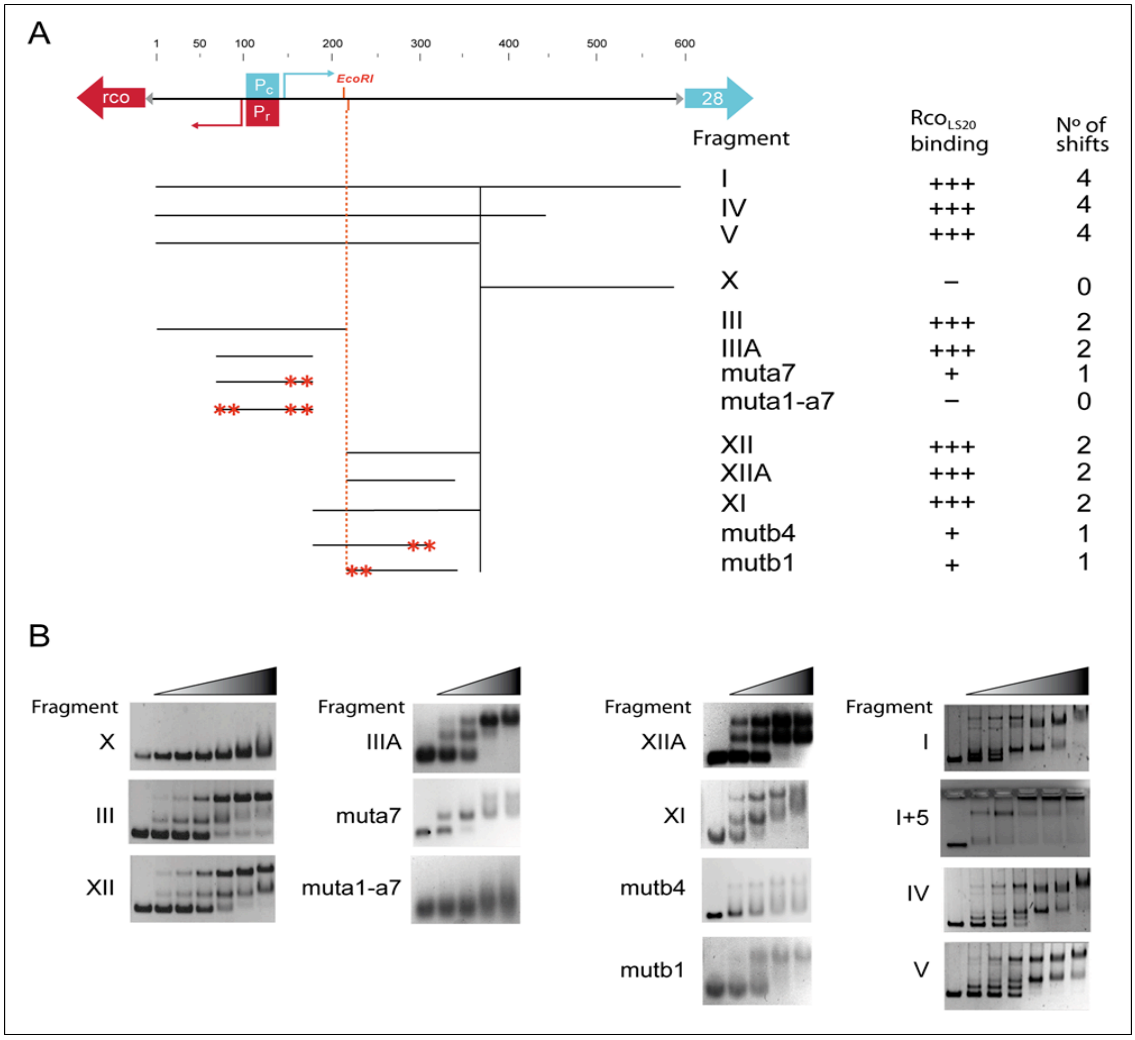
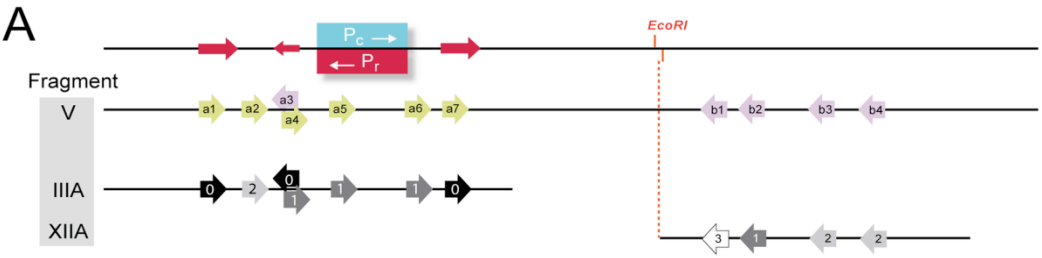


Figure R.4 Analysis of Rco_{LS20} binding sites in the *rco_{LS20} - gene 28* intergenic region by EMSA. **A**. Schematic representation of the DNA fragments used as probes. Top panel. A map of the 600 bp *rco_{LS20} -gene 28* intergenic region is shown on the top alongside with positions of characteristics in this region. The divergently oriented genes rco_{LS20} and 28 are indicated with a red and blue arrow, respectively. The triangles indicate ribosomal binding sites. The position of the P_c and the P_r promoter (boxed) and their transcription start sites (bent arrows) are indicated in blue and red, respectively. The position of the unique EcoRI site in this region is indicated. The different fragments subjected to EMSA are indicated; their names level of Rco_{LS20} binding and number of retarded species are indicated on the right. Asterisks indicate positions of introduced mutations. **B.** Overview of the results of EMSA employing the indicated fragments. Rco_{LS20} concentrations, two-fold dilution steps, ranged from 0.212-6.8 (6 lanes) or 0.95-7.3 µM (4 lanes). No shift was observed for an unrelated 550 bp fragment when incubated with 7.3 µM of Rco_{LS20}-$His_{(6)}$ (not shown).

A



B

Nº	Operator	Name	Strand	Sequence (5´-3´)		
1	II	a7	+	gtgaaagagt	**CAGTGAAA**	aaaatgcaga
2	II	a3	–	taaagtattt	**CAGTGAAA**	tttttaaccg
3	II	a1	+	tttttttact	**CAGTGAAA**	aaaaccggtt
4	II	a4	+	ttaaaaattt	**CA**C**TGAAA**	tactttacag
5	II	a5	+	aaatacttta	**CAGT**T**AAA**	aaaatgtctg
6	I	b4	–	ctcatacata	**CGGT**T**AAA**	ttaactggta
7	II	a2	+	tgaaaaaaac	**CGGT**T**AAA**	aatttcactg
8	II	a6	+	tctgttatct	T**AGTGAAA**	gagtcagtga
9	I	b2	–	tcagagacaa	**CAGTGAA**T	acgttggtta
10	I	b3	–	aattaactgg	T**AGT**T**AAA**	acttcagaga
11	I	b1	–	gtgaatacgt	TG**GT**T**AAA**	acaaaagttg

C

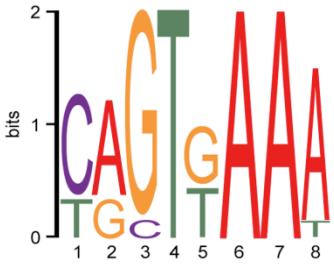


Figure R.5. **Identification of conserved motif constituting (part of) the Rco_{LS20} binding site**.
A. Schematic representation of the region corresponding to Fragment V, which encompasses promoters P_r/P_c and identified repeated motifs that form (part of) binding site for Rco_{LS20}. Top line. Promoter P_c and P_r and their transcriptional start sites (bent arrows) are indicated in blue and red, respectively. The position of the unique EcoRI site is indicated. A 13 bp long direct repeat (5´-TCAGTGAAAAAAA-3´) is indicated with leftward directed red arrows. The rightward-arrow indicates the position of the complementary 9 bp sequence 5´-TTTCACTGA-3´. Second line (Fragment V). Arrows indicate the position of the identified motifs a1-a7 and b1-b4. Motifs present on upper and lower strand are given in green and purple, respectively. Third and fourth line show identified motifs present on Fragment III_A and XII_A, respectively. Black, dark grey, grey, light grey and white indicate motifs identical to the consensus sequence or deviating at none, one, two and three positions, respectively. **B**. An alignment of the nucleotide sequences of the eleven identified motifs and their flanking sequences. Names according the nomenclature in "A" are given together with information on the strand and region. Sequences corresponding to the consensus sequence of the motif are given in white. **C**. A representation of the consensus motif generated by Weblogo(97). The size of each nucleotide corresponds to the frequency with which that nucleotide was observed in that position.

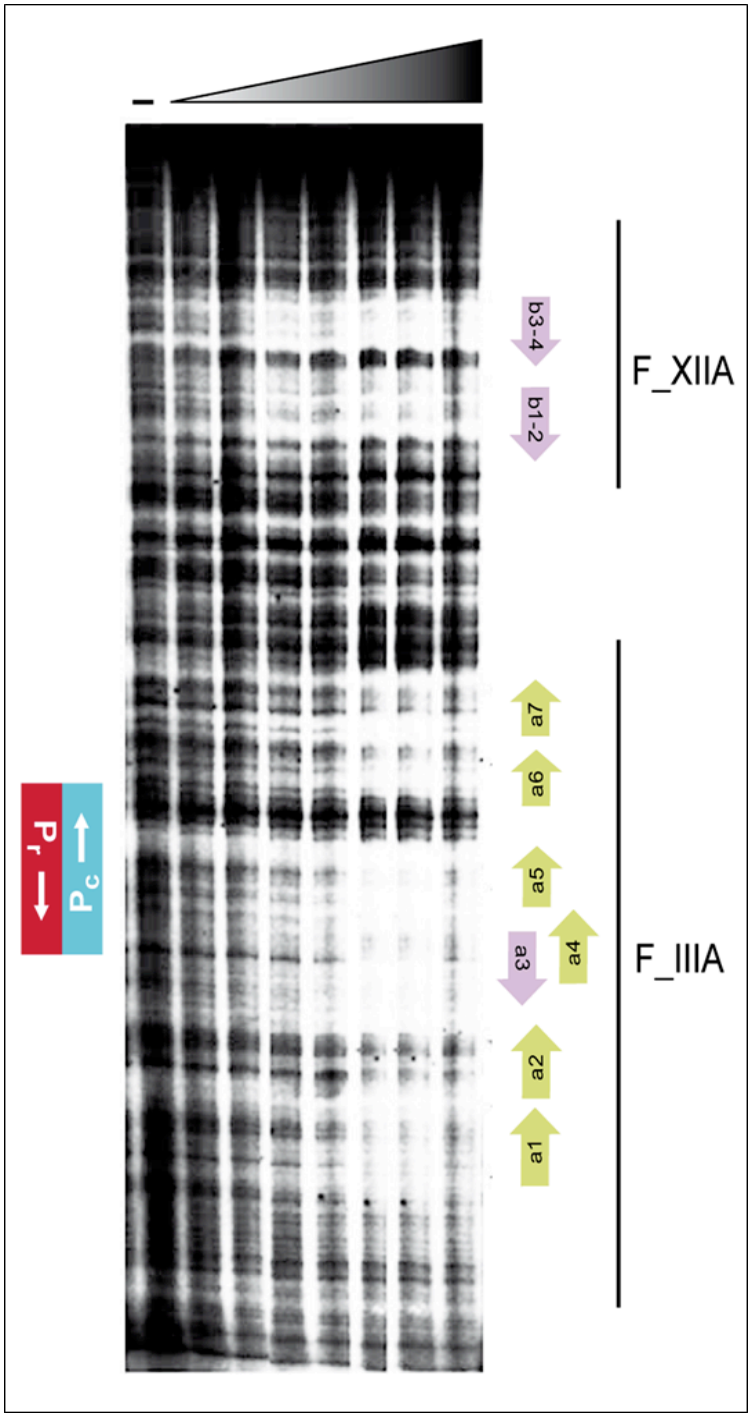


Figure R.6. **Footprint analyses of the binding of Rco_{LS20} to the *rco_{LS20}-gene 28* intergenic region.** Fragment V, end-labeled at the P_c template strand, was analyzed for binding of Rco_{LS20}-His. First lane (-) was not incubated with protein and the next lanes were incubated with Rco_{LS20}-His, two-fold dilution steps, ranged from 0.11 to 7.04 µM. The positions of the P_c and P_r promoters are indicated on the left. Bars on the right reflect the regions covered by Fragments IIIA (F III_A) and XIIA (F XII_A). Positions of motifs a1-a7 and b1-b4 are indicated with green or purple arrows at the right.

R.1.7 Evidences that proper regulation of the P_r/P_c promoters involves DNA looping

Operator O_I, -located at a distance of more than 75 bp from P_r/P_c-, is needed for proper regulation of these promoters. This and other data presented above, suggest that proper regulation of the P_r/P_c promoters involves DNA looping mediated by Rco_{LS20} bound to operators O_I and O_{II}. Due to the intrinsic stiffness of DNA, loops are generally larger than 90 bp because the curvature energy required to make smaller loops is too great, unless the DNA region separating the two operator sites is bent. Operators O_I and O_{II} are separated by only 75 bp. Several periodically spaced A/T tracts can result in formation of a static bent. The spacer region contains periodically spaced A/T tracts, and computer-assisted analysis predicts that the spacer region forms a static curve (see **Fig.R.7**). These data prompted us to perform circular permutation assays. Thus, three overlapping fragments of identical size (314 bp) were generated in which the predicted static curve is located at different positions (see **Fig. R.8.A**). As expected, these fragments migrated to identical positions when run on a 2% agarose gel (**Fig. R.8.B**). However, when run on a native 8% PAA gel the fragments migrated differently and all of them run slower than expected for their size, with the fragment containing the predicted bent in the middle of the fragment migrating slowest (**Fig. R.8.C**). These results show that the 75 bp spacer contains a static bent. If Rco_{LS20}-mediated DNA looping occurs then it is expected (i) that Rco_{LS20} will form oligomers thereby creating a DNA binding unit able to bind simultaneously to O_I and O_{II}, and (ii) that the two operators are in phase such that the Rco_{LS20} binding sites have a spatial orientation that is optimal for Rco_{LS20} binding. We tested both predictions. The oligomerization state of Rco_{LS20} was studied by two complementary analytical ultracentrifugation approaches (**Fig. R.9**). In sedimentation velocity experiments, Rco_{LS20} was observed as a single species with an experimental sedimentation coefficient of 3.8 S. This value corrected to standard conditions ($s20,w$ = 4.1S) was compatible with an elongated protein tetramer (**Fig. R.9.A**). To confirm this result, sedimentation equilibration experiments were carried out within the concentration

range from 10 to 30 µM. The calculated average molecular mass obtained was 85,200 Da ± 1,700, which corresponds to the tetrameric form of Rco_{LS20} (**Fig. R.9.B)**.

To test if a specific phasing between O_I and O_{II} is important for Rco_{LS20} to carry out its regulatory role we constructed a derivative of Fragment I, I+5, in which we enlarged the spacer half a helical turn by inserting 5 bp and cloned this fragment in front of *lacZ* (see **Fig. R.2.B**). Next, we tested the responsiveness of promoter P_c to Rco_{LS20} using strains containing either Fragment F_I_c or F_I_c +5 fused to *lacZ*. As expected, Rco_{LS20}, which was provided *in trans* by pLS20cat, efficiently repressed promoter P_c when *lacZ* was fused to Fragment I_c (strain PKS8). Promoter P_c was not efficiently repressed by Rco_{LS20} however, when *lacZ* was fused to Fragment I_c+5 (strain GR191). Thus, colonies of pLS20cat-harboring cells were blue when grown on Xgal-containing plates (see **Fig. R.10**). These results show that enlarging the distance between O_I and O_{II} with half a helical turn destroys proper regulation of promoter P_c by Rco_{LS20}. Besides affecting the phasing, the 5 bp insertion might also affect the static curvature of the spacer region. Regardless whether the loss of Rco_{LS20}-mediated regulation is due to phasing and/or altered curvature, the results provide compelling evidence that Rco_{LS20} mediates its regulatory effect through DNA looping. Next, we analyzed by EMSA if the 5 bp insertion between operators O_I and O_{II} affects Rco_{LS20} binding. As described above, even in the presence of the highest Rco_{LS20} concentration applied, the various DNA fragments entered the gel migrating to distinct positions indicating multiple intramolecular Rco_{LS20} binding events. Interestingly, however, whereas Fragment I+5 entered the gel at low Rco_{LS20} concentrations, most of the DNA did not enter the gel at medium or high Rco_{LS20} concentrations (**Fig R.3**). One possible explanation is that dephasing between the two operators allows Rco_{LS20} to bind intermolecularly resulting in the formation of high molecular weight nucleoprotein complexes that do not enter the gel. Together, these results support the view that the phasing between O_I and O_{II} is crucial for proper Rco_{LS20}-mediated regulation of transcription.

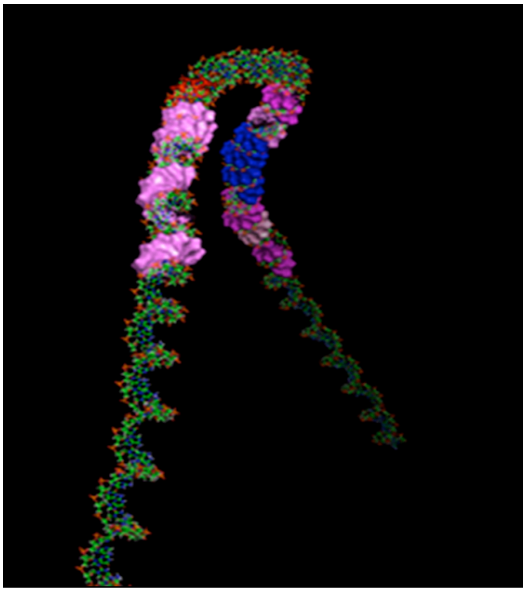


Figure R.7. **The 75 bp region separating operators O_I and O_{II} is predicted to contain a static bent.** The global 3D structure of a 256 bp DNA region encompassing operators O_I and O_{II} was predicted according the dinucleotide wedge mode using the online webpage http://www.lfd.uci.edu/~gohlke/dnacurve/. For clarity, sequences corresponding to promoters P_c/P_r and motifs in operators O_I and O_{II} are presented as space filling. Positions of the promoters and motifs are given in blue and purple respectively

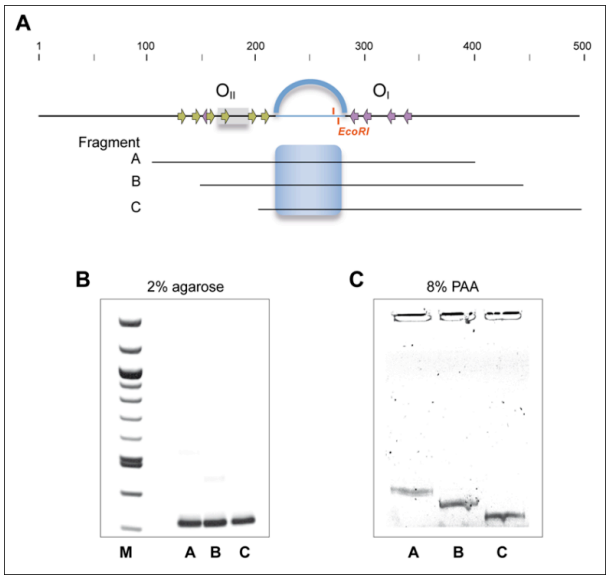


Figure R.8. .**The 75 bp spacer region separating operators O_I and O_{II} contains a static bent**.A. Schematic representation of the region encompassing operators O_I and O_{II} and the regions corresponding to PCR amplified fragments subjected to 2% agarose (B) and 8% native PAA . Position of the P_c/P_r promoters and the Rco_{LS20} binding motifs within the operators are indicated with grey rectangles and arrows, respectively. The 75 bp region separating O_I and O_{II} is shown as an interrupted line and the position of the unique EcoRI site is given. The predicted curvature in this region is represented by the arc above the line and by a shaded region to indicate the position in fragments A-C analyzed. Fragments A-C were run on 2% agarose on 8% native PAA gel followed by ethidium bromide staining.

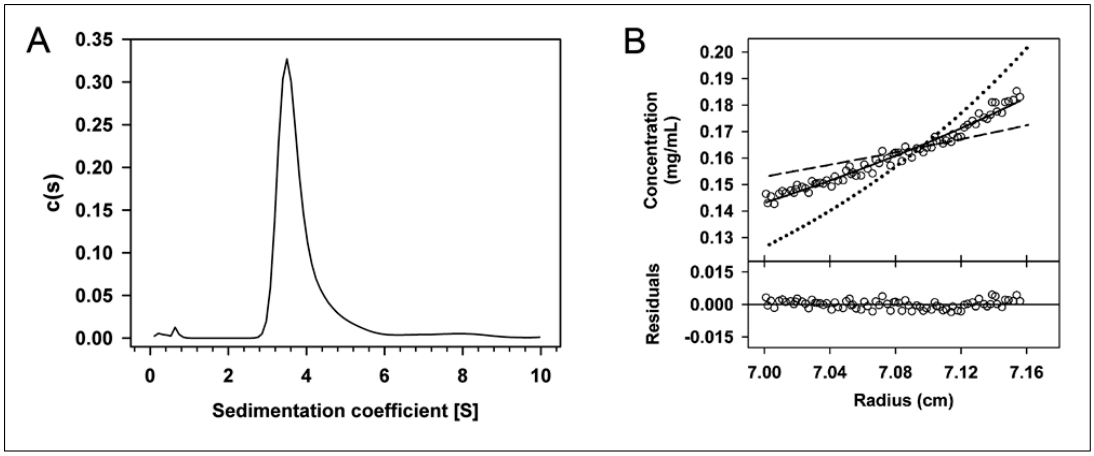


Figure R.9. .**Rco_{LS20} forms tetramers in solution.**
Sedimentation velocity studies. A): coefficient distributions (c(s)) recovered for Rco_{LS20} at 3.75 µM. B) Molecular masses of the distinct rco_{LS20} species determined by sedimentation equilibrium. Open circles represent the experimental gradients (UV absorption of the sample versus radial position in the centrifuge cell) at sedimentation equilibrium for Rco_{LS20} at 3.75 µM. Solid line is the best fit gradient for single sedimenting tetrameric species. OR: solid line shows the theoretical gradient of the tetrameric form of Rco_{LS20}. Lower panel, lower part. The residuals between experimental velocity data and best fit distribution model (c(s) were evenly distribute*d*

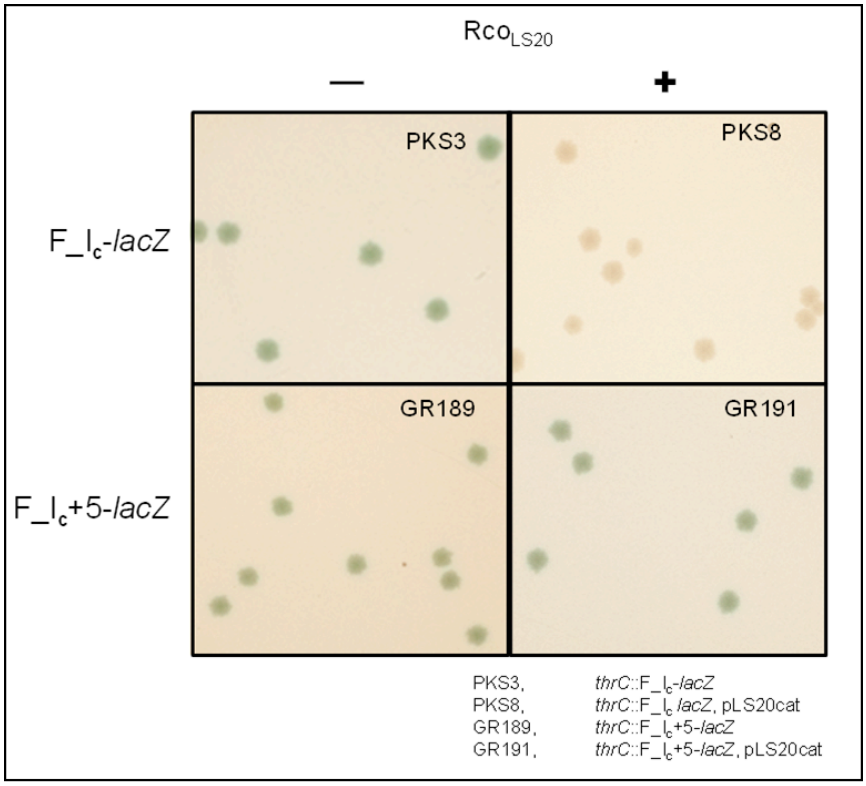


Figure R.10: **Enlarging the distance between operators O_I and O_{II} with half a helical turn affects Rco_{LS20}-mediated inhibition of promoter P_c.** Strains containing F_I$_c$ and F_I$_c$+5 fused to *lac*Z (PKS3 and GR189, respectively) and their derivatives harboring pLS20*cat* (PKS8 and GR191, respectively) were spread on Xgal-containing LB agar plates and photographed after 24 hours incubation at 37 °C

Chapter II:

An essential step of conjugation-formation of the relaxosome during the transfer of plasmid pLS20 to the recipient cell

R.2.1 Identification of pLS20*cat* genes 56, 57 and 58 encoding putative relaxosome proteins and are essential for conjugation

One of the initial steps of the conjugation process is the formation of the relaxosome, in which a Relaxase, -sometimes with the help of auxiliary proteins-, introduces a strand- and site-specific nick in the DNA strand, named T-strand, that is destined to be transferred to the recipient cell. The DNA region containing the nicking site is called " origin of transfer" or *ori*T.

Most information available about relaxases and *ori*T´s comes from studies on conjugative plasmids of G- bacteria or from mobilizable rolling-circle replicating plasmids (32, 43). As for other topics related with conjugation, little is known about components of the relaxosome of conjugative plasmids from G+ bacteria. Besides its intrinsic scientific interest, we decided to identify and characterize the relaxosome components of pLS20 principally for the following two additional reasons. First, this information will provide insights between the relatedness of relaxosomes between conjugative elements replicating in G+ and G- bacteria. Second, the information is crucial to design conjugation-based tools to genetically manipulate industrially and clinically relevant G+ bacteria that are reluctant to genetic modification by other methods.

As explained in more detail in the introduction, conjugative elements generally contain a single relaxase-encoding gene, and a single *ori*T located often upstream. In some cases, two or more genes encoding auxiliary relaxosome proteins are located upstream the relaxase gene.

We initiated these studies with *in silico* analyses to identify the putative relaxase gene(s) of pLS20. A combination of BLASTP and psi-BLAST analyses of all putative proteins encoded by pLS20*cat* against protein sequences present in the non-redundant database revealed that the putative pLS20*cat* gene *58* (410 codons) may encode a relaxase. An alignment of the pLS20*cat* deduced P58 protein sequence with (putative)

relaxases is presented in **Figure R.11.** This alignment shows that the gene product of pLS20*cat* gene *58* shares the highest level of homology with (putative) relaxases encoded by conjugative plasmids of other G+ bacteria. A more extensive description of the relatedness between the putative pLS20-encoded relaxase and other relaxases is given below.

Figure R.11: **Comparison of the protein 58 with other relaxases belonging to the family MOB$_{(MG)}$** The conserved motifs have been marked, W(X_4) H(X_2) T(X_3) HXH(X_4) E(X_4) R in the motifIII, is the representative motif of these types of Mob proteins.

pLS20*cat gene 58* is located within the conjugation operon(90). A schematic representation of the pLS20 region-encompassing *gene 58* is presented in **Figure R.12**.

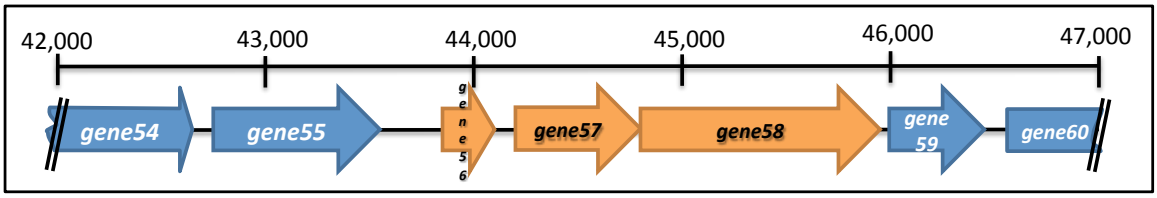


Figure R.12:**A schemmatic overview of pLS20*cat* region encompassing the genes *56-58*.** The probable *mob* genes are marked in orange, while the rest of the genes are marked in blue.

The proteins encoded by pLS20*cat* genes *54* and *55* show homology with VirB1 conjugation proteins and peptidases, respectively (not shown), and are therefore not predicted to form part of the relaxosome. The putative small genes *56* and *57*, however, might encode proteins involved in formation of a relaxosome. This assumption is made on the fact that they are both predicted to encode a DNA binding (see **Discussion Chapter II**). In the case that genes *56* to *58* would encode the relaxosome proteins it is expected that deletion of these genes would eradicate the ability of pLS20*cat* to conjugate, but not affect its replication. To test this hypothesis we constructed a derivative of pLS20*cat*, pLS20Δ*56-58*, in which genes *56* to *58* are deleted. The deletion of this region did not affect replication of pLS20 (not shown). Next, *B. subtilis* 168 cells containing pLS20*cat* (strain PKS11) or pLS20Δ*56-58* (strain GR149) were used as donor cells to determine their maximum conjugation efficiency level using a standard protocol (see **Materials & Methods**). The efficiency of conjugation observed for pLS20*cat* was in the range of 10^{-3}, which is similar to values reported before under these condition (90). However, no transconjugants were observed for pLS20Δ*56-58*.

These results suggest that at least one of the genes *56* to *58* is required for conjugation, however, the loss of conjugation observed for pLS20Δ*56-58* might also be due to deletion of the *oriT*, which is essential for conjugation, or because of the deletion introduced in pLS20Δ*56-58* causes polar effects on downstream genes in the conjugation operon. To rule out these possibilities, we constructed strain GR150 containing a cassette at its *amy*E-locus in which genes *56-58* are placed under the control of the IPTG-inducible P_{spank} promoter. A derivative of this strain harbouring pLS20Δ*56-58*, strain GR150, was used to measure the conjugation efficiencies in the absence or presence of 1 mM IPTG. Whereas no transconjugants were obtained when

GR150 cells were grown in the absence of IPTG, conjugation efficiencies similar to those observed with pLS20*cat* (in the range of 10^{-3}) were observed when cells were grown in the presence of IPTG. These results show that (i) pLS20Δ*56-58* contains a functional *ori*T, and (ii) one or more of the genes *56*, *57* and *58* is required for conjugation. Moreover, the results demonstrate that expression of genes *56-58* from an ectopic locus can complement their deletion on plasmid pLS20Δ*56-58*, resulting in full restoration of conjugation. This feature was then used to study the essentiality of each of these three genes for conjugation. Thus, we constructed strain GR206 that contains genes *56* and *57* (but not *58*) under the control of the P_{spank} promoter at the *amyE* locus and harbouring pLS20Δ*56-58*. No transconjugants were obtained when this strain was used as donor in conjugation experiments, independent whether the cells grew in the presence or absence of IPTG. These results show that pLS20*cat* gene *58* is required for conjugation. Taking into account the similarity of the deduced protein sequence with relaxase genes (see **Fig. R.11** and below), we conclude that pLS20*cat* gene *58* encodes the relaxase, which we name Rel_{LS20}.

Using similar strategies, evidence was obtained that also gene *56* and *57* are essential for conjugation. Thus, we constructed strain GR197 containing a cassette in which genes *57* and *58* are under the control of the P_{spank} promoter and harbouring pLS20Δ*56-58*. No transconjugants were obtained when this strain was used as donor in conjugation experiments in the absence or presence of IPTG.

To analyse gene *57*, we constructed strain GR200. This pLS20Δ*56-58*-harboring strain contains two cassettes; one in which gene *58* (*rel*$_{LS20}$) is under the control of the P_{spank} promoter, and another one in which gene *56* is under the control of the xylose-inducible P_{xyl} promoter. No transconjugants were observed when strain GR200 was used as donor in conjugation experiments in the absence or presence of both inductors. Together, these results provide strong evidence that all three genes are required for conjugation of pLS20.

Table T.3**: Conjugation efficiencies of pLS20*cat* and pLS20Δ*56-58*.**

Strain	**Genotype**	**Plasmid**	**Inductor**	**Conjugation efficiency ***
PKS11	168	pLS20*cat*	-	$8.3x10^{-3}$
GR149	168	pLS20Δ*56-58*	-	$< 10^{-7}$
GR150	168, *amyE*::P_{spank} *56-58*	pLS20Δ*56-58*	-	$< 10^{-7}$
			+	$1.39x10^{-3}$
GR206	168, *amyE*::P_{spank} *56-57*	pLS20Δ*56-58*	-	$< 10^{-7}$
			+	$< 10^{-7}$
GR197	168, *amyE*::P_{spank} *57-58*	pLS20Δ*56-58*	-	$< 10^{-7}$
			+	$< 10^{-7}$
GR200	168, *amyE*::P_{spank} *58, lacA*::P_{xyl}*-56*	pLS20Δ*56-58*	-	$< 10^{-7}$
			+	$< 10^{-7}$

: Conjugation efficiencies are calculated as* transconjugants/donor. *Conjugation efficiencies are the mean value of at least three independent experiments. When indicated, the inductor IPTG was added at a final concentration of 1 mM in the case of strains GR150, GR206 and GR197. In the case of GR200 IPTG (1 mM) and xylose (2%) was added when indicated.

R.2.2 Identification of the origin of transfer *ori*T of pLS20*cat*

Functional *ori*T loci generally correspond to a region of several hundred bp (104). An *ori*T region contains the nicking site (*nic*) at which the relaxase cleaves the phosphodiester bond of a specific dinucleotide. The region 5´ of the *nic* site, which is the last portion of the ssDNA that enters the recipient cell, is referred to as the upstream or trailing region of a transferring stand. The region 3´ of the *nic* site is referred to as the downstream or leading region (41). Proteins forming the relaxosome complex usually can function in *trans*, and the results presented above show that this also accounts for the relaxosome proteins of pLS20cat. The presence of the functional *ori*T locus of pLS20cat on a non-mobilizable compatible plasmid is therefore expected to be mobilized by pLS20cat. We used this functional approach to identify the *ori*T locus of pLS20cat, which we will name *ori*T_{LS20} henceforth. Thus, we engineered the

*ori*T screening vector pUCTA2501 and screened derivatives containing regions of pLS20 for their ability to be mobilized by pLS20cat. The *E. coli*/*B. subtilis* shuttle vector pUCTA2501 contains the replication functions of the cryptic *B. subtilis* rolling-circle plasmid pTA1015 (35) as well as the erythromycin resistance gene of pE194. As expected, control experiments showed that pUCTA2501 is not mobilized by pLS20*cat*.

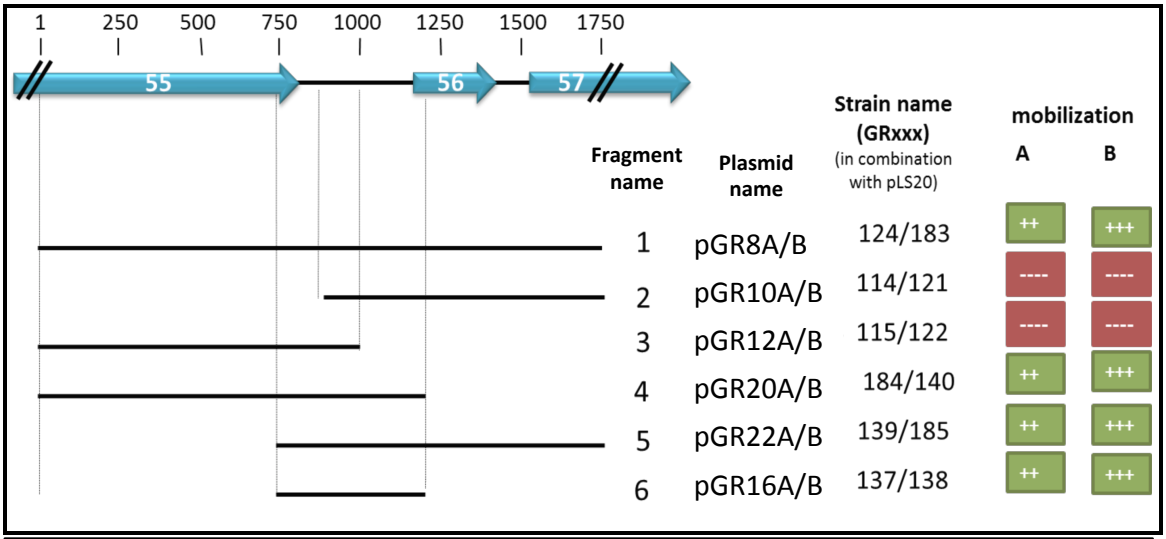


Figure R.12: **A schematic representation of the regions cloned in a non-mobilizable plasmid pUCTA2501, to visualize mobilization of the fragments by pLS20*cat*.** The strain GR105 (only containing the pUCTA2501 and pLS20*cat*) produced no transconjugants. The symbol + represents the transfer of the plasmid to the recipient cells and –-- represent no transfer of the RCR-plasmid, +++>++.

The *oriT* regions on other conjugative plasmids are often located upstream of the relaxase or accessory genes (34, 104). We therefore started analyzing whether *ori*T_{LS20} could be present on a 1.75 kb fragment, named Fragment 1, which encompasses the 3´-region of gene *55* till the approximately the middle of gene *57* (see **Figure R.11** for a schematic representation). Thus, Fragment 1 was cloned in pUCTA2501 and the mobilization efficiencies of the resulting plasmids pGR8A and pGR8B were determined ("A" and "B" correspond to the orientation of the cloned fragment tested). As shown **Table T.4** and schematically in **Figure R.11** contrary to pUCTA2501, pGR8A/B were efficiently mobilized by pLS20cat indicating that *ori*T_{LS20} is located on the 1.75 kb region present in these plasmids. To delineate the *ori*T_{LS20} region further several

internal regions of Fragment 1 were cloned onto pUCTA2501 and the mobilization efficiencies of the corresponding derivatives was determined. Using this approach, *ori*T_{LS20} was delineated to a 362 bp long region (Fragment 6 present in vectors pGR16A/B) corresponding principally to the gene *55-56* intergenic region (see **Table T.4** and **Figure R.11**).

Table T.4. **Mobilization efficiencies of the *ori*T-screening vector pUCTA2501 and derivatives**

Strain	Plasmid §	Region cloned (bp)	Mobilization efficiency *
GR104	pUCTA2501		$< 10^{-7}$
GR124	pGR8A	1739	$3.33X10^{-5}$
GR183	pGR8B	1739	$1.25\ X10^{-4}$
GR114	pGR10A	849	$-< 10^{-7}$
GR121	pGR10B	849	$< 10^{-7}$
GR115	pGR12A	949	$< 10^{-7}$
GR122	pGR12B	949	$< 10^{-7}$
GR184	pGR20A	1152	$4.5\ X10^{-5}$
GR140	pGR20B	1152	$1.71\ X10^{-4}$
GR139	pGR22A	949	$6.8\ X10^{-5}$
GR185	pGR22B	949	$1.17X\ X10^{-4}$
GR137	pGR16A	362	$2\ X10^{-5}$
GR138	pGR16B	362	$1.46\ X10^{-4}$

*§: Besides the plasmid mentioned, all strains contained pLS20cat. *: Mobilization efficiencies are calculated as Em-resistant* transconjugants/donor and using strain PS110 ($spec^R$) as recipient strain. Mobilization *efficiencies are the mean value of at least three independent experiments. The "A" and "B" extensions of the pUCTA2501 derivatives reflect different orientations of the same insert.*

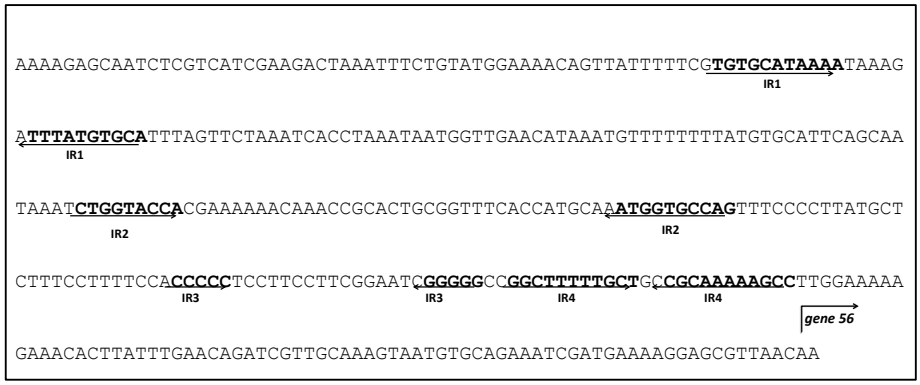


Fig. R.14: **Nucleotide sequence of the *ori*T region of plasmid pLS20.** The inverted repeats/stem loops have been indicated by arrows and the start of *gene 56*.

The features of the *ori*T region, include high AT rich region, allowing negative super coiling of the DNA region, further a higher presence of direct and indirect repetitions and also intrinsic bends allowing the proteins to bind altering the *ori*T structure locally. The two plots above indicate the higher AT content further the bendability of the

region (48). Extensive nucleotide sequence similarities have been found between related plasmids like F, IncP or IncQ group. In many cases, the *sso* of RC-type replication systems, showed similarities with the *nic* sites of several conjugative plasmids (48, 105). Based on the available consensus *nic* sites, we have postulated a *nic* site for the plasmid pLS20*cat* would lie in the sequence 5'-aatggtgccagtt-3'.

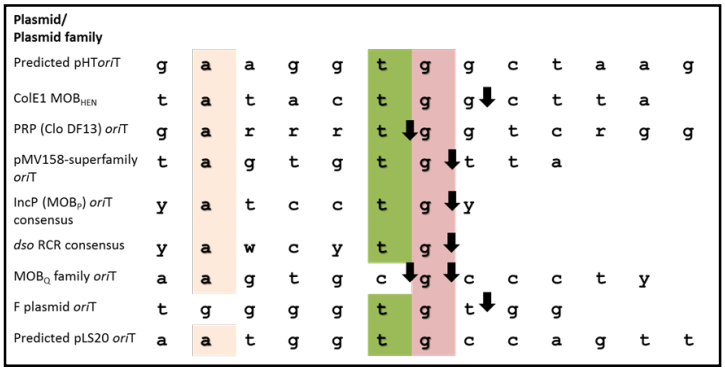


Figure R.15: **A comparative study of the consensus *nic* sites of various plasmid families, along with the predicted *nic* site of pLS20cat**. The conserved A, four positions upstream the nicking site has been marked with a light orange colour, the consensus T and G have been highlighted in light green colour and light red colour respectively. The black arrows specify the nicking site. The relaxase of the plasmid pHT*oriT* was found to be similar to the protein 58 of plasmid pLS20. Further, other plasmids are representative plasmids of the various groups.

5. DISCUSSION

Chapter I:

D.1 A complex genetic switch involving overlapping divergent promoters and DNA looping regulates expression of conjugation genes of a Gram-positive plasmid

Conjugation is a complex and energy consuming process, involving the generation and transfer of ssDNA, synthesis and assembly of a sophisticated type IV secretion system, and establishment of specific contacts with the recipient cell. Hence, the process of conjugation and expression of the genes involved are strictly controlled. Analysis of the regulation of conjugation genes present on ICEs in bacteria and those on plasmids of G- bacteria indeed indicates that this is the case (for review(23, 52)). However, with the exception of some enterococcal plasmids (106), far less is known about regulation of conjugation genes present on plasmids of G+ bacteria. In our previous studies, we have sequenced and annotated the *B. subtilis* plasmid pLS20*cat* and identified a large conjugation operon. We have also identified rco_{LS20} as the gene encoding the master regulator of conjugation, Rap_{LS20} as the anti-repressor required to activate the conjugation genes, and we showed that the activity of Rap_{LS20} is in turn regulated by the signaling peptide Phr^*_{LS20}. In this study, we analyzed the underlying molecular mechanism of how the pLS20 conjugation genes are regulated. The results obtained provide compelling evidence that a unique genetic switch, which is composed of at least three intertwined layers, controls activation of the pLS20 conjugation genes. A scheme of the three layers is shown in **Figure D.1**.

One of the levels results from the relative positioning of the main conjugation promoter, P_c, and the divergently oriented promoter P_r, driving expression of the rco_{LS20} gene (**Fig. D.1A**). We identified the conjugation promoter P_c and showed that it is a relatively strong promoter, which is repressed by the master regulator of conjugation Rco_{LS20}. Importantly, the position of promoter P_c coincides, or at least partially overlaps, with the divergently oriented weak P_r promoter. It has been demonstrated that an RNA polymerase can bind only to one of two overlapping promoters(107, 108). Thus, in the special configuration of overlapping promoters the RNA polymerase may itself act as a transcriptional regulator. Recently, Bendtsen *et al* (109) described theoretical scenarios backed up by experimental data that overlapping promoters indeed can result in a transcriptional switch, provided that they have different activities in the absence of the regulatory protein, combined with a regulator

that has a strong differential effect on the regulation of both promoters. This is exactly the case for the P_c/P_r promoter pair; in the absence of the regulator promoter P_c is several hundred folds stronger than P_r, and the presence of the regulator strongly represses the Pc promoter while activating the P_r promoter.

The second level of regulation contributing to the genetic switch concerns the multiple roles that Rco_{LS20} plays in the P_c/P_r regulation (**Fig. D.1B**). We showed that, on the one hand, Rco_{LS20} activates transcription of its own weak promoter, P_r, thereby generating a self-sustaining positive feedback loop. On the other hand, RcoLS20 functions simultaneously as an efficient repressor of the P_c promoter. The dual effect that RcoLS20 has on Pc and Pr maintains conjugation effectively in the "OFF" state. We also showed that the level of *rcoLS20* induction from an inducible promoter required for efficient repression of the Pc promoter was about ten-fold lower than that required for maximum auto-activation of the Pr promoter. These differential effects of RcoLS20 on repressing and activating the P_c/P_r promoters will also contribute towards maintaining conjugation stably in the "OFF" state during conditions when conjugation is not activated. Interestingly, we found that at elevated concentrations Rco_{LS20} inhibits its own transcription. This negative auto-regulation probably functions to keep Rco_{LS20} within a low concentration range in order to respond accurately to the anti-repressor Rap_{LS20} to activate the conjugation genes.

The triple effects Rco_{LS20} has on the regulation of the P_c/P_r promoters will also play an important role when Rap_{LS20} induces the system to switch to the "ON" state. In addition to relieving repression of the strong conjugation P_c promoter, this will simultaneously annihilate autostimulation of the P_r promoter, preventing further synthesis of RcoLS20, which in turn will contribute in pushing and maintaining conjugation in the "ON" state. A third level contributing to the genetic switch to activate the conjugation genes involves the DNA looping mediated by simultaneous binding of Rco_{LS20} to operators O_I and O_{II} (**Fig. D.1C**). DNA looping mediated by a transcriptional regulator has been reported for several other regulatory systems in

prokaryotes and their analyses has revealed that several features are conserved and necessary for DNA looping to occur (for review, see (110)). Our results show that the properties of Rco_{LS20} and the DNA in the P_c/P_r region complies with the necessary features for Rco_{LS20}-mediated loop formation. First, using different techniques, we show that Rco_{LS20}, -predicted to contain a helix-turn-helix DNA binding motif in its N-terminal region-(90), is a DNA binding protein and that it binds specifically to two operators, O_I and O_{II}. Second, operator O_I, which is located more than 85 bp away from promoters P_c and P_r, is required for efficient regulation of both promoters. Third, RcoLS20 binds cooperatively to both operators. Fourth, dephasing the positions of the two operators by inserting 5 bp in the spacer region destroys proper regulation of the conjugation genes. And fifth, we showed that Rco_{LS20} forms tetramers in solution. This will create a unit containing multiple DNA binding motifs, facilitating cooperative binding to multiple sites within the two operators. The DNA loop in the P_c/P_r region of pLS20 is characterized by a small spacer region that separates Rco_{LS20} operators O_I and O_{II}. The spacer length can be used to classify DNA loops into two categories: short or energetic loops, or long or entropic ones. Due to intrinsic stiffness and torsional rigidity of the DNA, loop formation is normally unfavorable for those with spacer lengths shorter than the DNA persistence length (approximately 150 bp), because the curvature energy required forming such small loops becomes too great. For such short loops to occur specific features like intrinsic static bending or binding of an additional protein inducing bending are required. In the case of pLS20, in which the operators O_I and O_{II} are separated by only 75 bp, we show that the spacer region contains a static bent.

The first experimental demonstration that a DNA loop can play a crucial role in transcriptional regulation was reported for the *E. coli ara* operon in 1984 (111). Since then, transcriptional regulator-mediated DNA looping has been demonstrated to regulate some other operons (for reviews see, (110, 112–114)). Although bio-informatical analyses suggests that DNA looping may be involved in the regulation of many genes and operons (110), the actual number of transcriptional systems for which

DNA looping has been conclusively demonstrated is remarkably low. In the case of plasmids, reports demonstrating DNA looping systems is limited to only few cases. One of these includes regulation of initiation of DNA replication at the beta origin of the *E. coli* R6K plasmid (115); and in the case of *Enterococcus faecalis* plasmid pCF10 it has been proposed that regulation of its conjugation system involves DNA looping mediated by the pheromone-responsive transcriptional regulator PrgX (for review see, (66)).

What are the benefits of DNA looping in general and for the regulation of the conjugation genes of pLS20 in particular? A major consequence of DNA looping is that it results in a high local concentration of the transcriptional regulator at the right place, which would increase its specificity and affinity [for recent review see, (116)]. Often, -and RcoLS20 is not an exception-, transcriptional regulators are produced in limited amounts per cell. Low numbers of regulators enhance the possibility of transcriptional fluctuations between individual cells within a population. In addition, the intrinsic stochasticity of transcription, - also referred to as noise-, affects the temporal effectiveness of transcriptional regulation; again this is especially prominent when the number of regulatory proteins involved is low. Recent evidences indicate that DNA looping contributes importantly to control temporal transcriptional noise, as well as dampening transcriptional fluctuations between cells within a population (112). Thus, DNA looping contributes to the tight regulation of promoters especially when levels of transcriptional regulators are low by diminishing stochastic fluctuations in transcription. For some differentiation processes, cell-to-cell or stochastic variability in levels of transcriptional regulators form the basis for activation of these processes, resulting in different behavior of genetically identical cells within a population (117–119). Examples of these processes are the formation of persister cells, development of natural genetic competence, spore formation and swimming/chaining. It is believed that such a bet-hedging strategy is beneficial for the fitness of the species because there will always be some cells that are prepared to cope with a deteriorating environmental condition that may arise in the near future. However, for other

processes, there may not be such an advantage and it would then be important to tightly repress the process at times when conditions for that process are not apt. Conjugation probably is such a process because there is no benefit in activating the conjugation genes when there is no recipient present to receive the plasmid. The fact that the efficiency of pLS20 transfer during growth conditions antithetic to conjugation is below the detection limit (at least six orders of magnitude lower than those observed during optimal conjugation conditions) strongly indicates that conjugation genes are tightly repressed under such conditions. However, the tight repression of conjugation should not compromise the ability of rapidly switching to high expression of the conjugation genes when appropriate conditions occur. In pLS20 this is achieved by the constellation of DNA looping combined with autoregulated expression of Rco_{LS20} and overlapping divergent promoters of different strength.

In summary, in this work we have provided evidence that regulation of the conjugation genes present on pLS20 is based on a unique genetic switch that combines, three levels of control, all reported for the first time for conjugation. These include (i) overlapping divergent promoters of different strengths, (ii) auto-stimulation and -repression of the weak P_r promoter by the transcriptional regulator at low and elevated concentrations, respectively, combined with simultaneous repression of the divergent strong conjugation promoter, and (iii) DNA looping mediated by binding of RcoLS20 regulator to two operators separated by a short loop. Most likely, the combination of these different layers causes tight repression of the main conjugation promoter P_c when conditions for conjugation are not optimal, while allowing the system to switch rapidly to high expression of the conjugation genes when appropriate conditions occur.

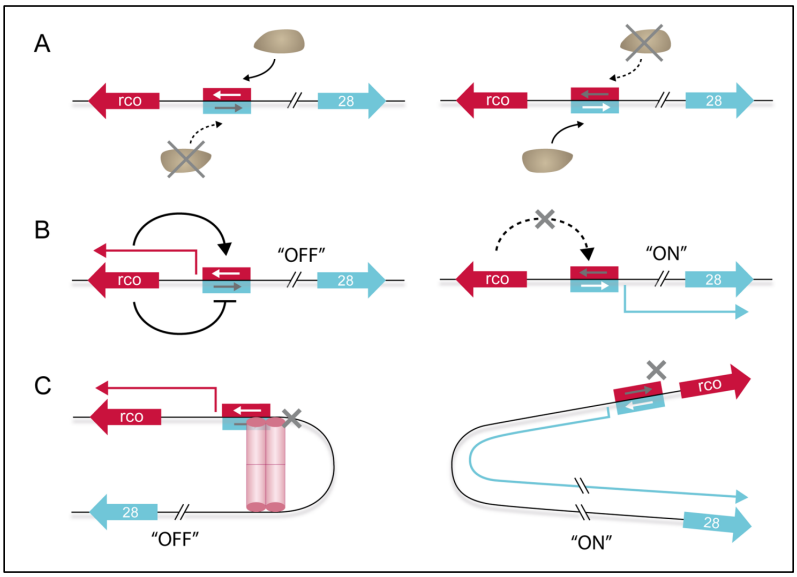


Figure D.1 **Model of the different layers contributing to the genetic switch controlling expression of the pLS20 conjugation genes.**

A. RNA polymerase acts itself as a switch because it is unable to bind simultaneously to both of the two overlapping and divergently oriented promoters. Consequently, RNA polymerase (the brown ellipse shaped form) binds only one promoter at a time resulting in transcription of only the gene(s) controlled by this promoter.

B. Rco_{LS20} generates a self-sustaining positive feedback loop by activating transcription from its own promoter (Pr) (left panel). This, combined with the simultaneous repression of the divergent conjugation promoter (P_c), results in conjugation being maintained effectively in the "OFF" state. Relief of Rco_{LS20}-mediated repression of the Pc promoter results in activation of the conjugation genes (right panel). In addition, this interrupts the auto-stimulation of the Pr promoter, preventing further synthesis of Rco_{LS20}, which in turn will contribute in pushing and maintaining conjugation in the "ON" state. The negative auto-regulatory loop of Rco_{LS20} that probably functions to keep Rco_{LS20} within a low concentration range (see text) is not presented.

C. DNA looping results in a high local concentration of Rco_{LS20}, increasing specificity and affinity that dampens transcriptional fluctuations between and within individual cells (left panel). This would contribute to tight repression of the P_c promoter, keeping conjugation in the "OFF" state under conditions antithetic to conjugation without compromising the ability to switching rapidly to a high expression state (i.e. "ON", right panel) of the conjugation genes when appropriate conditions occur. *rco_{LS20}* and gene *28*, -the first gene of the conjugation operon-, are indicated with large red and blue arrows, respectively. The same coloring scheme is used for the corresponding promoters (rectangular) and transcripts (thin broken arrows). Activation and repression of transcription are indicated with continuous black lines ending in an arrow and a "T" shape, respectively. The red cylindrical structures, which may reflect one or two $RcoL_{S20}$ tetramers, represent the Rco_{LS20} oligomer mediating DNA looping.

Chapter II:

D.2 An essential step of conjugation-formation of the relaxosome during the transfer of plasmid pLS20 to the recipient cell

Conjugative DNA strand transfer is a highly conserved mechanism for the unidirectional transfer of genetic information among bacteria of the same species (7), from one species to another and across kingdoms (120). The inevitable processes of conjugative DNA transfer are the formation of a stable mating pair; establishment of cell-to-cell contact is required for the physical transfer of ssDNA from the donor to the recipient. The process is followed by the production of a site- and strand-specific nick at a locus on the conjugative plasmid called *nic* which lies within the origin of transfer *ori*T. The unwinding of the duplex DNA molecule to provide the ssDNA, which leads to the transfer of the recipient cells this process, is coupled with CPs proteins and the T4SS, which allows the transfer. This process has been thoroughly studied in G- bacteria, the few well studied plasmids include, F-plasmid, R388 and RP4 the former two isolated from *E. coli* and the latter from *Pseudomonas aeruginosa* (41).

Origin of transfer region of a conjugative plasmid are termed *cis*-acting site required for DNA transfer and they contain the region to be nicked and from which the strand is unwound and transported to the recipient (48). We have shown by sequence alignment that the *ori*T of pLS20 has the necessary consensus sequences to be 5'-aatggtgccagtt-3' and experimentally we need to determine the *nic* site, which "lies" within this proposed *ori*T region (see **Figure R.12**). *oriT* region and *nic* sites within for conjugative plasmids can be crudely divided into two groups, the F-family and the P-family of conjugative plasmids (see **Figure D.2**, (41)). *nic* sites of mobilizable plasmids have been found to be conserved, one such type is the RSF1010-*ori*T family and the conserved nucleotides were deduced to be 5´-NcgtNtaAgtGCGCcCTta-3´(48), others include ColE1 or pMV158-superfamily.

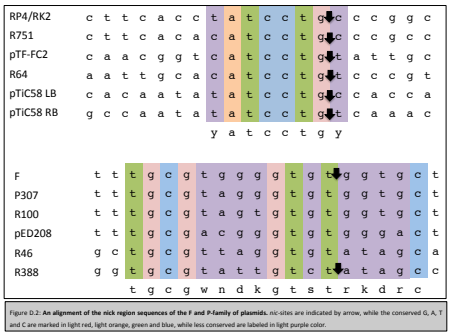


Figure D.2: **An alignment of the nick region sequences of the F and P-family of plasmids.** nic-sites are indicated by arrow, while the conserved G, A, T and C are marked in light red, light orange, green and blue, while less conserved are labeled in light purple color.

We have initial data showing that the proteins p56, p57 and p58 derived from genes *56*, *57* and *58* of the plasmid pLS20. Plasmid pLS20Δ*56-58* is unable to conjugate, unless genes *56* to *58* are ectopically produced. Thus, providing initial evidence of the association of proteins at the *ori*T region.

A *psi*-blast of the protein p56 showed us, that p56 is similar with 31% identity to a CopG-like protein encoded on the genome of *Clostridium sordellii* VPI 9048 (a G+ organism-gene H476_030). *Clostridium sordellii* is a toxigenic pathogen, although rare, infection with *C. sordellii* is a dangerous and life threatening (121, 122). Two toxins, a lethal and a hemorrhagic (that antigenically and pathophysiologically appear similar to *Clostridium difficile* toxins B and A, respectively) are responsible for the infection (122). Further, the region adjacent to gene H476_030 was found to consist of genes related to the T4SS genes, a hallmark of MGEs, thus, it might be involved in the transfer of PIs. The amino acid sequence of protein p57 was found to be 23% similar to ribbon-helix-helix type of protein coded on the genome of *Clostridium difficile* P49 (EQJ94812.1). *Clostridium difficile* 630 consists of 11% of genome consist of mobile elements. However, the complete sequence of *Clostridium difficile* P49 is still not available but, it reinforces that conjugative elements play an important role in transfer of pathogenic genomic islands encoding for virulence factors.

The main characteristics of RHH proteins include a) the N-terminal β-strand b) Loop α_1–α_2 and c) the hydrophobic core. N-terminal β-strand consists of amino acids with alternating hydrophilic-hydrophobic side chains, between the positions 2 and 7. The domain core consists of the hydrophobic amino acids at odd-numbered positions (3, 5 and 7), and hydrophilic containing side chain-containing amino acids pointing away, to permit the contact with the DNA strand. Usually arginine or lysine, are found at positions 2 and 4. A hydrophobic amino-acid side chain contacting DNA bases might provide some specificity, for example by allowing van der Waals interactions with the methyl group of thymine bases. The second feature of RHH proteins is the presence of a conserved G-X-S/T/N motif in the loop between helix α_1

and helix α_2. This loop is important for the correct positioning of the two loops. Third feature of a hydrophobic core consist of four positions 15, 27, 31 and 35 are conserved-hydrophobic amino acids in helix α_1 and helix α_2. The side chains of these hydrophobic amino acids with the hydrophobic amino acids at 3, 5 and 7 from the N-terminal β-strand, form most of the hydrophobic core of the RHH_2 domain (see review (123–126)).

A comparison of protein p56 with the prototype of the sub-family of CopG-like proteins of the ribbon-helix-helix family of proteins i.e CopG_pMV158 has been carried out (127) (see **Figure D.3**), p56, may be called $CopG_{LS20}$. CopG of plasmid pMV158 has been observed to hinder the binding of the RNA polymerase and also binds a 13-bp pseudosymmetric primary DNA recognition element (126, 127). Thus, p56 "might" also have some role in enhancing the activity of the relaxase. Although, no such characteristic features were determined for p57, but based on blast and *in-vivo* results, it can assumed that it might also be related with the processing of the *ori*T.

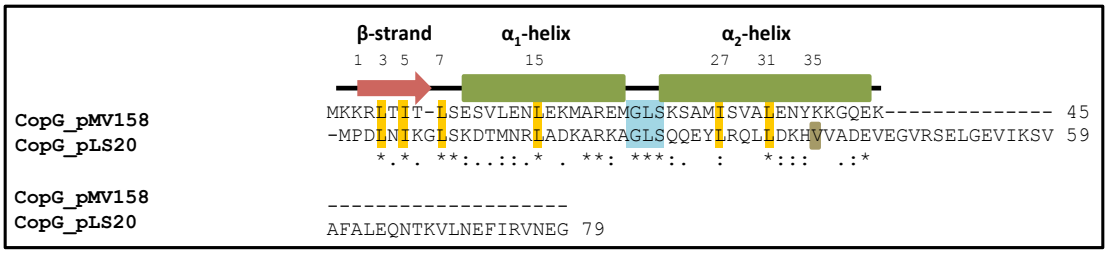


Figure D.3: **A comparison of the RHH domain of CopG_LS20 with a classical RHH containing CopG protein P13920 encoded on the plasmid pMV158.**Alongwith the predicted the important/conserved amino acids making the RHH domains, which includes the hydrophobic amino acids marked in yellow and the loop-turn is marked in blue, further a conserved amino acid is seen in CopG_LS20 marked in green (which is not conserved in CopG_MV58).

RHH-type proteins have been associated with plasmids like R388, R1 and F found in G- bacteria. TrwA of plasmid R388 belongs to the Arc repressor superfamily and consists of two domains, N-terminal DNA binding and C-terminal tetramerization domain. By structural studies, it was found that TrwA polar amino acids Q8, R10, S12, are directly involved in DNA-binding, in a manner highly conserved in RHH proteins (128). Further, it was found that TrwA binding to the *trwABC* operon, led to

the repression of the expression of the *trwB* and *trwC*. In addition, TrwA enhances the activity of the relaxase TrwC *in vitro*, in co-relation to a 10^5 increase in the frequency of conjugation *in-vivo* (128, 129).

Other RHH proteins related with the DNA processing include TraY and TraM of F-plasmid. TraY (also identified in R1 plasmid) is an RHH-type of essential protein, which along with integration host factor (IHF) loads the TraI (relaxase) on the DNA, while TraM is a non-essential protein stimulating the relaxosome-mediated cleavage at *ori*T through an interaction with TraI (130–132). Thus, RHH-type of proteins seem to be functional during DNA processing at the *ori*T. This data can be a used in the future to determine the binding site of each of proteins p56, p57 and Rel_{LS20}, may have independent/overlapping sites of binding to modulate the action of the relaxase.

Conjugative plasmids have been classified on the basis of their relaxases. Relaxases are a diverse group of proteins, of varying sizes. The N-terminal domain, show the nicking activity, while the C-terminal domain can have differing functions like DNA-helicase, DNA-primase or other functions. Additionally, the N-terminal consists of three motifs, motif I contains the catalytic Tyr residue involved in the cleavage-joining activity, further is temporally covalently linked to the 5'-terminus of the *nic* site, motif II involved in DNA-protein contacts through the 3'-end of the *nic* region and a Ser residue is usually present and motif III consists three conserved His residues and is known as 3H motif. This motif is proposed to be involved in the nucleophilic activity of Tyr residue in Motif I with the requirement of Mg^{2+} ions and direct activation of the active Tyr (32, 34, 41, 43).

Mob proteins are classified into six families (43), based on their N-terminal amino acid sequence and the Tyrosine motif. Families, which only contain one Tyr in their active sites, are MOB_P, MOB_Q and MOB_V, while family MOB_F contains two tyrosine-moieties in its active site. MOB_{HEN} is part of the MOB_P family of relaxases. Also MOB_P

family is understudied family due to paucity of sequences, however the family seems to overlap with MOB_Q and MOB_V families. Finally, families MOB_H and MOB_C catalyze the transesterification reaction by an unknown mechanism (43).

Incidentally, Rel_{LS20} does not appear to fit into the above-mentioned families of relaxases. On the basis of the conserved motif, it is similar to ORF34 of an Enterococcal-conjugative plasmid pHTβ. ORF34 of the plasmid pHTβ along with the relaxases/nickases of other plasmids like plasmids pOX2 from *B. anthracis*, pHTβ-*Enterococci*, p19 from *Bacillus* spps, p576 from *B. pumilus*, pLI100 from *Listeria* and pGL5 from *Lactococcus*, (all isolated from G+ organisms) have been compared with Rel_{LS20}. They constitute a new family of Mob proteins known as the $MOB_{(MG)}$ family with a conserved motif; $W(X_4)$ $H(X_2)$ $T(X_3)$ $HXH(X_4)$ $E(X_4)R$ in the motif III of the protein domains (105). Thus, a ClustalW alignment of $Rel_{LS20}$with above-mentioned plasmids has been shown in **Figure R.11**, we have marked the conserved motif (W (X_4) H (X_2) T (X_3) HXH (X_4) E (X_4) R), to highlight that Rel_{LS20} of plasmid pLS20 forms a representative of this ¨new¨ family (see **Figure R. 11)**.

The *mob* region consists of a gene encoding for the relaxase (gene *58*), which is preceded by two genes, one coding for a Cop-like (RHH-type) protein and another RHH type of protein respectively. A clustal omega alignment of the proteins upstream the “relaxases” of the above-mentioned plasmids was performed, and they seem to have a certain degree of similarity (see **Figure D.5** and **D.6**). The various features of RHH type proteins have been highlighted, though, pairwise identity of RHH sequences is low and no amino-acid positions are completely conserved, several motifs have been marked based on inspection of the sequence and structural alignments. Further, *in-vivo* and *vitro* assays have to be performed, in order to determine the functions of the gene products. Also, more bio-informatical analysis has to be carried out to discover new plasmids (/and conjugative elements), which could form part of this novel family of Mob proteins.

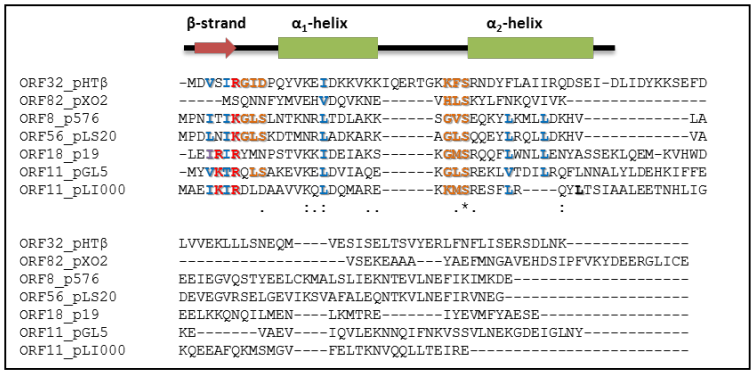


Figure D.5: **A clustal omega comparison of the protein sequences of p56 with the accessory proteins of the plasmid belonging to the group MOB$_{(MG)}$.** The ¨conserved¨ hydrophobic amino acids have been marked in blue and positively charged amino acids are marked in red and the amino acids which may form the turn have been marked in orange.

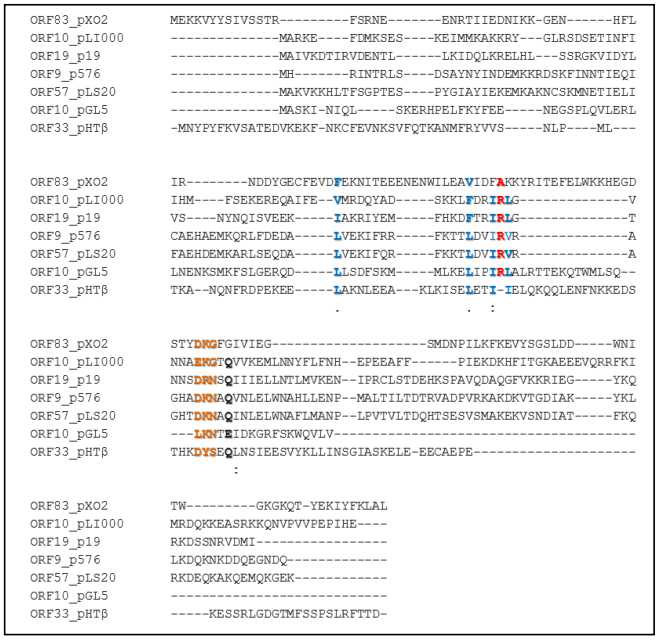


Figure D.6: **A clustal omega comparison of the protein sequences of p57 with the accessory proteins of the plasmid belonging to the group MOB$_{(MG)}$.** The ¨conserved¨ hydrophobic amino acids have been marked in blue and positively charged amino acids are marked in red and the amino acids which may form the turn have been marked in orange.

In summary, we have carried out a genetic analysis involved in DNA processing of the *Bacillus* plasmid pLS20, and found that the relaxase (Rel$_{LS20}$), carrying out the nicking reaction belongs to a new family of Mob proteins known as MOB$_{(MG)}$, the *ori*T region of plasmid pLS20 is upstream the gene *56*, which along with gene *57*, encode for DNA-binding proteins which may be auxiliary proteins mediating the processing at the *ori*T.

6. CONCLUSIONS

1) The promoter of conjugation P_c is under the control of the master regulator Rco_{LS20}, a DNA-binding protein of the helix-turn-helix type.
2) Using various experiments (namely β-gal activity, primer extension assay and RNAseq), it was found that Promoter P_c is located at an unusually large distance upstream found in the intergenic region between *gene 28* and rco_{LS20} of *gene 28* with a sequence 5´- ttaaaaatttcactgaaatac-**TTtACA**-gttaaaaaaatgtc-TGt**TATctT**-3´.
3) Rco_{LS20} was found to be a triple-functional protein, at low concentrations it is activates its promoter P_r while (the second function) repressing the promoter of conjugation P_c and at higher concentrations represses its own promoter.
4) Promoter of repression P_r was determined (5´-aaGAtA- 17bp -TgTAAa-3´), it was also deduced to be a weak promoter and was localized to be overlapping with the promoter of conjugation P_c.
5) Rco_{LS20} represses the promoter of conjugation P_c by binding to two operator O_I and O_{II} sites one located more than 85 bps downstream P_c- shown by *in-vivo* and *in-vitro* methods. Based on sequence similarities and other results, the probable motifs of binding have a consensus sequence of CAGTGAAA, with one to three deviations.
6) By de-phasing the operator sites O_I and O_{II} (by the addition of 5bp) and also the presence of static curvature between the two sites, we found evidences that proper regulation of the P_r/P_c promoters involves DNA looping.
7) *In-silico* and experimental evidence to identify the proteins/genes involved in the formation of the relaxasome, the genes were found to be *56*, *57* and *58* of the plasmid pLS20*cat*.
8) Identification of the origin of transfer *ori*T of the transfer gene complex of pLS20*cat* based on the sequence similarity can be proposed to be 5´-aatggtgccagtt-3´ found upstream the gene 56 of the plasmid pLS20*cat*.

6. CONCLUSIÓNES

1. El promotor de la conjugación P_c está controlado por el regulador maestro Rco_{LS20}, una proteína de unión a ADN de tipo hélice-giro-hélice.
2. Utilizando varios experimentos (Actividad β-gal, ensayos de extensión de cebador y RNAseq), descubrimos que el promotor P_c está localizado a una distancia inusualmente lejana hacia 5', en la región intergénica entre el *gen 28* y el rco_{LS20} del *gen 28*, con una secuencia 5´- ttaaaaatttcactgaaatac-**TTtACA**-gttaaaaaaatgtc-TGt**TATctT**-3´.
3. Descubrimos que Rco_{LS20} es una proteína trifuncional, (primera función) a niveles de inducción bajos por un locus ectópico activa a su promotor P_r mientras (segunda función) reprime al promotor de la conjugación P_c y (tercera función) a niveles de inducción más elevados reprime a su propio promotor.
4. El Promotor de represión P_r fue determinado (5´-aaGAtA- 17pb -TgTAAa-3´), también se dedujo que se trata de un promotor débil y que solapa con el promotor de la conjugación P_c.
5. Rco_{LS20} reprime al promotor de la conjugación P_c uniéndose a dos operadores O_I y O_{II}, uno localizado a más de 85 pb por debajo del P_c- lo que fue demostrado por métodos tanto *in vivo* como *in vitro*. Basándose en similitudes de secuencia y otros resultados, los motivos de unión probables tienen una secuencia consenso CAGTGAAA, con una a tres desviaciones.
6. Mediante desfase de los sitios de los operadores O_I y O_{II} (añadiendo 5pb) y también por la presencia de curvatura estática entre los dos sitios, encontramos evidencias de que la regulación apropiada de los promotores P_r/P_c implica a un lazo de ADN.
7. Tras evidencias *in silico* y experimentales para identificar a las proteínas/genes implicados en la formación del relaxosoma, las proteínas encontradas fueron p56, p57 y p58 del plásmido pLS20*cat*. La proteína p58, es una relaxasa, que necesita de proteínas auxiliares para llevar a cabo la actividad de corte de la hebra.

8. Basándonos en la similitud de secuencia se puede proponer que el sitio *nic* en la región de transferencia del plásmido pLS20*cat* está dentro de la secuencia 5´-aatggtgccagtt-3 encontrada a 5’ del *gen 56* del plásmido pLS20*cat*.

8. BIBLIOGRAPHY

1. Frost, L. S., Leplae, R., Summers, A. O., and Toussaint, A. (2005) Mobile genetic elements: the agents of open source evolution. *Nat. Rev. Microbiol.* **3**, 722–732

2. Stokes, H. W., and Gillings, M. R. (2011) Gene flow, mobile genetic elements and the recruitment of antibiotic resistance genes into Gram-negative pathogens. *FEMS Microbiol. Rev.* **35**, 790–819

3. Thomas, C. M., and Nielsen, K. M. (2005) Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat. Rev. Microbiol.* **3**, 711–21 [online] http://www.ncbi.nlm.nih.gov/pubmed/16138099 (Accessed February 21, 2014).

4. Spellberg, B., Guidos, R., Gilbert, D., Bradley, J., Boucher, H. W., Scheld, W. M., Bartlett, J. G., and Edwards, J. (2008) The epidemic of antibiotic-resistant infections: a call to action for the medical community from the Infectious Diseases Society of America. *Clin. Infect. Dis.* **46**, 155–164

5. Alekshun, M. N., and Levy, S. B. (2007) Molecular Mechanisms of Antibacterial Multidrug Resistance. *Cell* **128**, 1037–1050

6. Wright, G. D. (2007) The antibiotic resistome: the nexus of chemical and genetic diversity. *Nat. Rev. Microbiol.* **5**, 175–186

7. De la Cruz, F., and Davies, J. (2000) Horizontal gene transfer and the origin of species: lessons from bacteria. *Trends Microbiol.* **8**, 128–33 [online] http://www.ncbi.nlm.nih.gov/pubmed/10707066.

8. Collier, R. J. (2001) Understanding the mode of action of diphtheria toxin: A perspective on progress during the 20th century. *Toxicon* **39**, 1793–1803

9. Kolstø, A.-B., Tourasse, N. J., and Økstad, O. A. (2009) What sets Bacillus anthracis apart from other Bacillus species? *Annu. Rev. Microbiol.* **63**, 451–476

10. Lawrence, J. G. (2005) Horizontal and Vertical Gene Transfer : The Life History of Pathogens. **12**

11. Lawrence, J. G., and Ochman, H. (1998) Molecular archaeology of the Escherichia coli genome. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 9413–9417

12. Griffith, F. (1928) Pn eumococcal Types. **XXVII**

13. Lorenz, M. G., and Wackernagel, W. (1994) Bacterial gene transfer by natural genetic transformation in the environment. *Microbiol. Rev.* **58**, 563–602

14. Zafra, O., Lamprecht-Grandío, M., de Figueras, C. G., and González-Pastor, J. E. (2012) Extracellular DNA Release by Undomesticated Bacillus subtilis Is Regulated by Early Competence. *PLoS One* **7**

15. Cohan, F. M., Roberts, M. S., and King, E. C. (1991) The Potential for Genetic Exchange by Transformation within a Natural Population of Bacillus subtilis. *Evolution (N. Y).* **45**, 1393–1421 [online] http://www.jstor.org/stable/2409888.

16. Chen, I., and Dubnau, D. (2004) DNA uptake during bacterial transformation. *Nat. Rev. Microbiol.* **2**, 241–249

17. Exchange, G., and Dna, M. Chapter 2 Genetic Exchange in Bacteria and the Modular Structure of Mobile DNA Elements. , 34–78

18. Maier, B., Chen, I., Dubnau, D., and Sheetz, M. P. (2004) DNA transport into Bacillus subtilis requires proton motive force to generate large molecular forces. *Nat. Struct. Mol. Biol.* **11**, 643–9 [online] http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3832999&tool=pmcentrez&rendertype=abstract (Accessed October 21, 2014).

19. Labrie, S. J., Samson, J. E., and Moineau, S. (2010) Bacteriophage resistance mechanisms. *Nat. Rev. Microbiol.* **8**, 317–327

20. Boyd, E. F., Davis, B. M., and Hochhut, B. (2001) Bacteriophage-bacteriophage interactions in the evolution of pathogenic bacteria. *Trends Microbiol.* **9**, 137–144

21. Osborn, a M., and Böltner, D. (2002) When phage, plasmids, and transposons collide: genomic islands, and conjugative- and mobilizable-transposons as a mosaic continuum. *Plasmid* **48**, 202–12 [online] http://www.ncbi.nlm.nih.gov/pubmed/12460536.

22. Oppenheim, A. B., Kobiler, O., Stavans, J., Court, D. L., and Adhya, S. (2005) Switches in bacteriophage lambda development. *Annu. Rev. Genet.* **39**, 409–29 [online] http://www.ncbi.nlm.nih.gov/pubmed/16285866 (Accessed January 22, 2014).

23. Frost, L. S., and Koraimann, G. (2010) Regulation of bacterial conjugation: balancing opportunity with adversity. *Future Microbiol.* **5**, 1057–1071

24. Davison, J. (1999) Genetic exchange between bacteria in the environment. *Plasmid* **42**, 73–91 [online] http://www.ncbi.nlm.nih.gov/pubmed/10489325.

25. Jain, R., Rivera, M. C., and Lake, J. a (1999) Horizontal gene transfer among genomes: the complexity hypothesis. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 3801–6 [online] http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=22375&tool=pmcentrez&rendertype=abstract.

26. Burrus, V., and Waldor, M. K. (2004) Shaping bacterial genomes with integrative and conjugative elements. *Res. Microbiol.* **155**, 376–86 [online] http://www.ncbi.nlm.nih.gov/pubmed/15207870 (Accessed February 28, 2014).

27. Amábile-Cuevas, C. F. (2012) Antibiotic Resistance: From Darwin to Lederberg to Keynes. *Microb. Drug Resist.*, 121009061121003

28. Sørensen, S. J., Bailey, M., Hansen, L. H., Kroer, N., and Wuertz, S. (2005) Studying plasmid horizontal transfer in situ: a critical review. *Nat. Rev. Microbiol.* **3**, 700–10 [online] http://www.ncbi.nlm.nih.gov/pubmed/16138098 (Accessed January 22, 2014).

29. Pogliano, J. (2002) Dynamic cellular location of bacterial plasmids. *Curr. Opin. Microbiol.* **5**, 586–590

30. Solar, G., Giraldo, R., Ruiz-echevarría, M. J., Espinosa, M., Díaz-orejas, R., and Di, N. (1998) Replication and Control of Circular Bacterial Plasmids Replication and Control of Circular Bacterial Plasmids. **62**

31. Khan, S. a (1997) Rolling-circle replication of bacterial plasmids. *Microbiol. Mol. Biol. Rev.* **61**, 442–55 [online] http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=232620&tool=pmcentrez&rendertype=abstract.

32. Smillie, C., Garcillán-Barcia, M. P., Francia, M. V., Rocha, E. P. C., and de la Cruz, F. (2010) Mobility of plasmids. *Microbiol. Mol. Biol. Rev.* **74**, 434–52 [online] http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2937521&tool=pmcentrez&rendertype=abstract (Accessed February 26, 2014).

33. Meijer, W. J., Venema, G., and Bron, S. (1995) Characterization of single strand origins of cryptic rolling-circle plasmids from Bacillus subtilis. *Nucleic Acids Res.* **23**, 612–619

34. Francia, M. V., Varsaki, A., Garcillán-Barcia, M. P., Latorre, A., Drainas, C., and De La Cruz, F. (2004) A classification scheme for mobilization regions of bacterial plasmids. *FEMS Microbiol. Rev.* **28**, 79–100

35. Meijer, W. J. J., Wisman, G. B. A., Terpstra, P., Thorsted, P. B., Thomas, C. M., Holsappel, S., Venema, G., and Bron, S. (1998) Rolling-circle plasmids from Bacillus subtilis: Complete nucleolide sequences and analyses of genes of pTA1015, pTA1040, pTA1050 and pTA1060, and comparisons with related plasmids from Gram-positive bacteria. *FEMS Microbiol. Rev.* **21**, 337–368

36. Meijer, W. J., de Jong, A., Bea, G., Wisman, A., Tjalsma, H., Venema, G., Bron, S., and van Dijl, J. M. (1995) The endogenous Bacillus subtilis (natto) plasmids pTA1015 and pTA1040 contain signal peptidase-encoding genes: identification of a new structural module on cryptic plasmids. *Mol. Microbiol.* **17**, 621–631

37. Koehler, T. M., and Thorne, C. B. (1987) Bacillus subtilis (natto) plasmid pLS20 mediates interspecies plasmid transfer. *J. Bacteriol.* **169**, 5271–5278

38. Farı, E., and Espinosa, M. (2000) Conjugal transfer of plasmid pMV158 : uncoupling of the pMV158 origin of transfer from the mobilization gene mobM , and modulation of pMV158 transfer in Escherichia coli mediated by IncP plasmids. **158**, 2259–2265

39. Firth, N., Ippen-ihler, K., and Skurray, R. A. Structure and Function of the F Factor and Mechanism of Conjugation.

40. Lawley, T. D., Klimke, W. A., Gubbins, M. J., and Frost, L. S. (2003) F factor conjugation is a true type IV secretion system. *FEMS Microbiol. Lett.* **224**, 1–15

41. Lanka, E., and Wilkins, B. M. (1995) DNA processing reactions in bacterial conjugation. *Annu. Rev. Biochem.* **64**, 141–169

42. Wong, J. J. W., Lu, J., and Glover, J. N. M. (2012) Relaxosome function and conjugation regulation in F-like plasmids - a structural biology perspective. *Mol. Microbiol.* **85**, 602–617

43. GarcillÃ¡n-Barcia, M. P., Francia, M. V., and de la Cruz, F. (2009) The diversity of conjugative relaxases and its application in plasmid classification. *FEMS Microbiol. Rev.* **33**, 657–687 [online] http://doi.wiley.com/10.1111/j.1574-6976.2009.00168.x (Accessed March 11, 2014).

44. Cascales, E., and Christie, P. J. (2003) The versatile bacterial type IV secretion systems. *Nat. Rev. Microbiol.* **1**, 137–49 [online] http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3873781&tool=pmcentrez&rendertype=abstract (Accessed March 19, 2014).

45. Christie, P. J. (2004) Type IV secretion: the Agrobacterium VirB/D4 and related conjugation systems. *Biochim. Biophys. Acta* **1694**, 219–34 [online] http://www.ncbi.nlm.nih.gov/pubmed/15546668 (Accessed February 4, 2014).

46. Ding, Z., Atmakuri, K., and Christie, P. J. (2003) The outs and ins of bacterial type IV secretion substrates. *Trends Microbiol.* **11**, 527–535 [online] http://linkinghub.elsevier.com/retrieve/pii/S0966842X03002609 (Accessed February 4, 2014).

47. Llosa, M., Gomis-Rüth, F. X., Coll, M., and De la Cruz, F. (2002) Bacterial conjugation: A two-step mechanism for DNA transport. *Mol. Microbiol.* **45**, 1–8

48. Grohmann, E., Muth, G., and Espinosa, M. (2003) Conjugative Plasmid Transfer in Gram-Positive Bacteria Conjugative Plasmid Transfer in Gram-Positive Bacteria. **67**

49. Cascales, E., and Christie, P. J. (2003) The versatile bacterial type IV secretion systems. *Nat. Rev. Microbiol.* **1**, 137–149

50. Baron, C., Llosa, M., Zhou, S., and Zambryski, P. C. (1997) VirB1, a component of the T-complex transfer machinery of Agrobacterium tumefaciens, is processed to a C-terminal secreted product, VirB1. *J. Bacteriol.* **179**, 1203–1210

51. Sagulenko, V., Sagulenko, E., Jakubowski, S., Spudich, E., and Christie, P. J. (2001) VirB7 lipoprotein is exocellular and associates with the Agrobacterium tumefaciens T pilus. *J. Bacteriol.* **183**, 3642–3651

52. Zatyka, M., and Thomas, C. M. (1998) Control of genes for conjugative transfer of plasmids and other mobile elements. *FEMS Microbiol. Rev.* **21**, 291–319

53. Blohm, D., and Goebel, W. (1978) Restriction map of the antibiotic resistance plasmid R1drd-19 and its derivatives pKN102 (R1drd-19B2) and R1drd-16 for the enzymes BamHI, HindIII, EcoRI and SalI. *Mol. Gen. Genet.* **167**, 119–127

54. Finlay, B. B., Frost, L. S., and Paranchych, W. (1986) traY allele from R100-1 . Origin of Transfer of IncF Plasmids and Nucleotide Sequences of the Type II oriT , traM , and tra Y Alleles from ColB4-K98 and the Type IV tra Y Allele from R100-1.

55. Ingram, L. C., Richmond, M. H., and Sykes, R. B. (1973) Molecular characterization of the R factors implicated in the carbenicillin resistance of a sequence of Pseudomonas aeruginosa strains isolated from burns. *Antimicrob. Agents Chemother.* **3**, 279–288

56. Komano, T., Funayama, N., Kim, S. R., and Nisioka, T. (1990) Transfer Region of IncII Plasmid R64 and Role of Shufflon in R64 Transfer. **172**

57. Zatyka, M., Jagura-Burdzy, G., and Thomas, C. M. (1997) Transcriptional and translational control of the genes for the mating pair formation apparatus of promiscuous IncP plasmids. *J. Bacteriol.* **179**, 7201–7209

58. Jagura-Burdzy, G., and Thomas, C. M. (1994) KorA protein of promiscuous plasmid RK2 controls a transcriptional switch between divergent operons for plasmid replication and conjugative transfer. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 10571–10575

59. Haase, J., Lurz, R., Grahn, A. M., Bamford, D. H., and Lanka, E. (1995) Bacterial conjugation mediated by plasmid RP4: RSF1010 mobilization, donor-specific phage propagation, and pilus production require the same Tra2 core components of a proposed DNA transport complex. *J. Bacteriol.* **177**, 4779–4791

60. Jagura-Burdzy, G., Khanim, F., Smith, C. A., and Thomas, C. M. (1992) Crosstalk between plasmid vegetative replication and conjugative transfer: repression of the trfA operon by trbA of broad host range plasmid RK2. *Nucleic Acids Res.* **20**, 3939–3944

61. Ziegelin, G., Furste, J. P., and Lanka, E. (1989) TraJ protein of plasmid RP4 binds to a 19-base pair invert sequence repetition within the transfer origin. *J. Biol. Chem.* **264**, 11989–11994

62. Pansegrau, W., Balzer, D., Kruft, V., Lurz, R., and Lanka, E. (1990) In vitro assembly of relaxosomes at the transfer origin of plasmid RP4. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 6555–6559

63. Ziegelin, G., Pansegrau, W., Lurz, R., and Lanka, E. (1992) TraK protein of conjugative plasmid RP4 forms a specialized nucleoprotein complex with the transfer origin. *J. Biol. Chem.* **267**, 17279–17286

64. Koraimann, G., and Wagner, M. a (2014) Social behavior and decision making in bacterial conjugation. *Front. Cell. Infect. Microbiol.* **4**, 54 [online] http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4010749&tool=pmcentrez&rendertype=abstract (Accessed November 17, 2014).

65. Singh, P. K., and Meijer, W. J. J. (2014) Diverse regulatory circuits for transfer of conjugative elements. *FEMS Microbiol. Lett.* [online] http://www.ncbi.nlm.nih.gov/pubmed/24995588.

66. Dunny, G. M. (2013) Enterococcal sex pheromones: signaling, social behavior, and evolution. *Annu. Rev. Genet.* **47**, 457–82 [online] http://www.ncbi.nlm.nih.gov/pubmed/24050179.

67. Dunny, G. M., and Leonard, B. A. B. (1995) MINIREVIEW Pheromone-Inducible Conjugation in Enterococcus faecalis : Interbacterial and Host-Parasite Chemical Communication. **177**, 871–876

68. Kao, S. M., Olmsted, S. B., Viksnins, A. S., Gallo, J. C., and Dunny, G. M. (1991) Molecular and genetic analysis of a region of plasmid pCF10 containing positive control genes and structural genes encoding surface proteins involved in pheromone-inducible conjugation in Enterococcus faecalis. *J. Bacteriol.* **173**, 7650–7664

69. Chandler, J. R., and Dunny, G. M. (2008) Characterization of the sequence specificity determinants required for processing and control of sex pheromone by the intramembrane protease Eep and the plasmid-encoded protein PrgY. *J. Bacteriol.* **190**, 1172–83 [online] http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2238190&tool=pmcentrez&rendertype=abstract (Accessed February 4, 2014).

70. Nakayama, J., Ruhfel, R. E., Dunny, G. M., Isogai, A., and Suzuki, A. (1994) The prgQ gene of the Enterococcus faecalis tetracycline resistance plasmid pCF10 encodes a peptide inhibitor, iCF10. *J. Bacteriol.* **176**, 7405–7408

71. Chandler, J. R., Flynn, A. R., Bryan, E. M., and Dunny, G. M. (2005) Specific control of endogenous cCF10 pheromone by a conserved domain of the pCF10-encoded regulatory protein PrgY in Enterococcus faecalis. *J. Bacteriol.* **187**, 4830–4843

72. Mori, M., Sakagami, Y., Ishii, Y., Isogai, A., Kitada, C., Fujino, M., Adsit, J. C., Dunny, G. M., and Suzuki, A. (1988) Structure of cCF10, a peptide sex pheromone which induces conjugative transfer of the Streptococcus faecalis tetracycline resistance plasmid, pCF10. *J. Biol. Chem.* **263**, 14574–14578

73. Leonard, B. A., Podbielski, A., Hedberg, P. J., and Dunny, G. M. (1996) Enterococcus faecalis pheromone binding protein, PrgZ, recruits a chromosomal oligopeptide permease system to import sex pheromone cCF10 for induction of conjugation. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 260–264

74. Chatterjee, A., Johnson, C. M., Shu, C.-C., Kaznessis, Y. N., Ramkrishna, D., Dunny, G. M., and Hu, W.-S. (2011) Convergent transcription confers a bistable switch in Enterococcus faecalis conjugation. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 9721–9726

75. Cook, L., Chatterjee, A., Barnes, A., Yarwood, J., Hu, W.-S., and Dunny, G. (2011) Biofilm growth alters regulation of conjugation by a bacterial pheromone. *Mol. Microbiol.* **81**, 1499–510 [online]

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3187857&tool=pmcentrez&rendertype=abstract (Accessed November 24, 2014).

76. Bensing, B. A., Manias, D. A., and Dunny, G. M. (1997) Pheromone cCF10 and plasmid pCF10-encoded regulatory molecules act post-transcriptionally to activate expression of downstream conjugation functions. *Mol. Microbiol.* **24**, 285–294

77. Chen, Y., Staddon, J. H., and Dunny, G. M. (2007) Specificity determinants of conjugative DNA processing in the Enterococcus faecalis plasmid pCF10 and the Lactococcus lactis plasmid pRS01. *Mol. Microbiol.* **63**, 1549–64 [online] http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2650854&tool=pmcentrez&rendertype=abstract (Accessed November 24, 2014).

78. Todar, K. (2005) in *Todar's online text book of bacteriology* p. 20

79. Kunst, F., Ogasawara, N., Moszer, I., Albertini, a M., Alloni, G., Azevedo, V., Bertero, M. G., Bessières, P., Bolotin, a, Borchert, S., Borriss, R., Boursier, L., Brans, a, Braun, M., Brignell, S. C., Bron, S., Brouillet, S., Bruschi, C. V, Caldwell, B., Capuano, V., Carter, N. M., Choi, S. K., Codani, J. J., Connerton, I. F., and Danchin, a (1997) The complete genome sequence of the gram-positive bacterium Bacillus subtilis. *Nature* **390**, 249–56 [online] http://www.ncbi.nlm.nih.gov/pubmed/9384377.

80. Sonenshein, A. L., Hoch, J. A., and Losick, R. (2002) *Bacillus subtilis and its closest relatives*, [online] http://books.google.com/books?id=26hYh4sE3hEC&pgis=1.

81. Pandey, a, and Palni, L. M. (1997) Bacillus species: the dominant bacteria of the rhizosphere of established tea bushes. *Microbiol. Res.* **152**, 359–65 [online] http://www.ncbi.nlm.nih.gov/pubmed/9468659 (Accessed March 14, 2014).

82. Bais, H. P., Fall, R., and Vivanco, J. M. (2004) Biocontrol of Bacillus subtilis against Infection of Arabidopsis Roots by Pseudomonas syringae Is Facilitated by Biofilm Formation and Surfactin. **134**, 307–319

83. Tanaka, T., and Koshikawa, T. (1977) Isolation and characterization of four types of plasmids from Bacillus subtilis (natto). Isolation and Characterization of Four Types of Plasmids from Bacillus subtilis (natto). **131**

84. Koehler, T. M., and Thorne, C. B. (1987) Bacillus subtilis (natto) plasmid pLS20 mediates interspecies plasmid transfer. *J. Bacteriol.* **169**, 5271–8 [online] http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=213936&tool=pmcentrez&rendertype=abstract.

85. Itaya, M., Sakaya, N., Matsunaga, S., Fujita, K., and Kaneko, S. (2006) Conjugational transfer kinetics of pLS20 between Bacillus subtilis in liquid medium. *Biosci. Biotechnol. Biochem.* **70**, 740–742

86. Meijer, W. J., de Boer, A. J., van Tongeren, S., Venema, G., and Bron, S. (1995) Characterization of the replication region of the Bacillus subtilis plasmid pLS20: a novel type of replicon. *Nucleic Acids Res.* **23**, 3214–3223

87. Derman, A. I., Nonejuie, P., Michel, B. C., Truong, B. D., Fujioka, A., Erb, M. L., and Pogliano, J. (2012) Alp7R regulates expression of the actin-like protein Alp7A in Bacillus subtilis. *J. Bacteriol.* **194**, 2715–24

88. Derman, A. I., Becker, E. C., Truong, B. D., Fujioka, A., Tucey, T. M., Erb, M. L., Patterson, P. C., and Pogliano, J. (2009) Phylogenetic analysis identifies many uncharacterized actin-like proteins (Alps) in bacteria: regulated polymerization, dynamic instability and treadmilling in Alp7A. *Mol. Microbiol.* **73**, 534–552

89. Singh, P. K., Ramachandran, G., Durán-Alcalde, L., Alonso, C., Wu, L. J., and Meijer, W. J. J. (2012) Inhibition of Bacillus subtilis natural competence by a native, conjugative plasmid-encoded comK repressor protein. *Environ. Microbiol.* **14**, 2812–25 [online] http://www.ncbi.nlm.nih.gov/pubmed/22779408 (Accessed March 24, 2014).

90. Singh, P. K., Ramachandran, G., Ramos-Ruiz, R., Peiró-Pastor, R., Abia, D., Wu, L. J., and Meijer, W. J. J. (2013) Mobility of the native Bacillus subtilis conjugative plasmid pLS20 is regulated by intercellular signaling. *PLoS Genet.* **9**, e1003892 [online] http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3814332&tool=pmcentrez&rendertype=abstract.

91. Sambrook, J., and Russell, D. W. (2001) Molecular Cloning - Sambrook & Russel - Vol. 1, 2, 3.pdf. *Hum. Mutat.* **18**, 1–2231

92. Singh, P. K., Ballestero-Beltrán, S., Ramachandran, G., and Meijer, W. J. J. (2010) Complete nucleotide sequence and determination of the replication region of the sporulation inhibiting plasmid p576 from Bacillus pumilus NRS576. *Res. Microbiol.* **161**, 772–782

93. Miller, J. H. (1972) Experiments in molecular genetics. *Cold Spring Harb. Lab. Press. Cold Spring Harb. NY* [online] http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=17416361.

94. Moreno, R., Fonseca, P., and Rojo, F. (2012) Two small RNAs, CrcY and CrcZ, act in concert to sequester the Crc global regulator in Pseudomonas putida, modulating catabolite repression. *Mol. Microbiol.* **83**, 24–40

95. Eckert, R. L. (2001) DNA sequencing by the chemical method. *Curr. Protoc. Mol. Biol.* **Chapter 7**, Unit7.5

96. Castilla-Llorente, V., Salas, M., and Meijer, W. J. J. (2008) kinC/D-mediated heterogeneous expression of spo0A during logarithmical growth in Bacillus subtilis is responsible for partial suppression of phi 29 development. *Mol. Microbiol.* **68**, 1406–17 [online] http://www.ncbi.nlm.nih.gov/pubmed/18410285 (Accessed February 4, 2014).

97. Bailey, T. L., Williams, N., Misleh, C., and Li, W. W. (2006) MEME: Discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res.* **34**

98. Liu, X., Brutlag, D. L., and Liu, J. S. (2001) BioProspector: discovering conserved DNA motifs in upstream regulatory regions of co-expressed genes. *Pac. Symp. Biocomput.*, 127–138

99. Bolshoy, a, McNamara, P., Harrington, R. E., and Trifonov, E. N. (1991) Curved DNA without A-A: experimental estimation of all 16 DNA wedge angles. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 2312–6 [online] http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=51221&tool=pmcentrez&rendertype=abstract.

100. Schuck, P. (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys. J.* **78**, 1606–1619

101. Laue, T. M., and Stafford, W. F. (1999) Modern applications of analytical ultracentrifugation. *Annu. Rev. Biophys. Biomol. Struct.* **28**, 75–100

102. Ucci, J. W., and Cole, J. L. (2004) Global analysis of non-specific protein-nucleic interactions by sedimentation equilibrium. *Biophys. Chem.* **108**, 127–140

103. Bailey, T. L., and Elkan, C. (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **2**, 28–36

104. De la Cruz, F., Frost, L. S., Meyer, R. J., and Zechner, E. L. (2010) Conjugative DNA metabolism in Gram-negative bacteria. *FEMS Microbiol. Rev.* **34**, 18–40 [online] http://www.ncbi.nlm.nih.gov/pubmed/19919603 (Accessed November 24, 2014).

105. Tomita, H., and Ike, Y. (2005) Genetic analysis of transfer-related regions of the vancomycin resistance Enterococcus conjugative plasmid pHT??: Identification of oriT and a putative relaxase gene. *J. Bacteriol.* **187**, 7727–7737

106. Shi, K., Brown, C. K., Gu, Z.-Y., Kozlowicz, B. K., Dunny, G. M., Ohlendorf, D. H., and Earhart, C. A. (2005) Structure of peptide sex pheromone receptor PrgX and PrgX/pheromone complexes and regulation of conjugation in Enterococcus faecalis. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 18596–18601

107. Wang, P., Yang, J., Lawley, B., and Pittard, A. J. (1997) Repression of the aroP gene of Escherichia coli involves activation of a divergent promoter. *J. Bacteriol.* **179**, 4213–4218

108. Yang, J., Wang, P., and Pittard, A. J. (1999) Mechanism of repression of the arop P2 promoter by the TyrR protein of Escherichia coli. *J. Bacteriol.* **181**, 6411–6418

109. Bendtsen, K. M., Erdossy, J., Csiszovszki, Z., Svenningsen, S. Lo, Sneppen, K., Krishna, S., and Semsey, S. (2011) Direct and indirect effects in the regulation of overlapping promoters. *Nucleic Acids Res.* **39**, 6879–6885

110. Cournac, A., and Plumbridge, J. (2013) DNA looping in prokaryotes: experimental and theoretical approaches. *J. Bacteriol.* **195**, 1109–19 [online] http://jb.asm.org/content/195/6/1109.full.

111. Dunn, T. M., Hahn, S., Ogden, S., and Schleif, R. F. (1984) An operator at -280 base pairs that is required for repression of araBAD operon promoter: addition of DNA helical turns between the operator and promoter cyclically hinders repression. *Proc. Natl. Acad. Sci. U. S. A.* **81**, 5017–20 [online] http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=391628&tool=pmcentrez&rendertype=abstract.

112. Saiz, L., and Vilar, J. M. (2006) DNA looping: the consequences and its control. *Curr. Opin. Struct. Biol.* **16**, 344–350

113. Dodd, I. B., Shearwin, K. E., and Egan, J. B. (2005) Revisited gene regulation in bacteriophage ?? *Curr. Opin. Genet. Dev.* **15**, 145–152

114. Lobell, R. B., and Schleif, R. F. (1990) DNA looping and unlooping by AraC protein. *Science* **250**, 528–532

115. Mukherjee, S., Erickson, H., and Bastia, D. (1988) Enhancer-origin interaction in plasmid R6K involves a DNA loop mediated by initiator protein. *Cell* **52**, 375–383

116. Oehler, S., and M??ller-Hill, B. (2010) High Local Concentration: A Fundamental Strategy of Life. *J. Mol. Biol.* **395**, 242–253

117. Veening, J.-W., Smits, W. K., and Kuipers, O. P. (2008) Bistability, epigenetics, and bet-hedging in bacteria. *Annu. Rev. Microbiol.* **62**, 193–210

118. Dubnau, D., and Losick, R. (2006) Bistability in bacteria. *Mol. Microbiol.* **61**, 564–572

119. Korobkova, E., Emonet, T., Vilar, J. M. G., Shimizu, T. S., and Cluzel, P. (2004) From molecular noise to behavioural variability in a single bacterium. *Nature* **428**, 574–578

120. Keeling, P. J., and Palmer, J. D. (2008) Horizontal gene transfer in eukaryotic evolution. *Nat. Rev. Genet.* **9**, 605–18 [online] http://www.ncbi.nlm.nih.gov/pubmed/18591983 (Accessed February 24, 2014).

121. Health, W. (2010) Fatal Clostridium sordellii Infections after Medical Abortions. , 1382–1383

122. Rørbye, C., Petersen, I. S., and Nilas, L. (2000) Postpartum Clostridium sordellii infection associated with fatal toxic shock syndrome. *Acta Obstet. Gynecol. Scand.* **79**, 1134–1135 [online] http://doi.wiley.com/10.1034/j.1600-0412.2000.0790121134.x.

123. Schreiter, E. R., and Drennan, C. L. (2007) Ribbon-helix-helix transcription factors: variations on a theme. *Nat. Rev. Microbiol.* **5**, 710–20 [online] http://www.ncbi.nlm.nih.gov/pubmed/17676053 (Accessed February 4, 2014).

124. Solar, G., Herna, A. M., Gomis-ru, F. X., Coll, M., and Espinosa, M. (2002) MINIREVIEW A Genetically Economical Family of Plasmid-Encoded Transcriptional Repressors Involved in Control of Plasmid Copy Number. **184**, 4943–4951

125. Popescu, A., Karpay, A., Israel, D. a, Peek, R. M., and Krezel, A. M. (2005) Helicobacter pylori protein HP0222 belongs to Arc/MetJ family of transcriptional regulators. *Proteins* **59**, 303–11 [online] http://www.ncbi.nlm.nih.gov/pubmed/15723352 (Accessed February 4, 2014).

126. Xavier Gomis-Rüth, F., Solà, M., Acebo, P., Párraga, A., Guasch, A., Eritja, R., González, A., Espinosa, M., Solar, G. Del, and Coll, M. (1998) The structure of plasmid-encoded transcriptional repressor CopG unliganded and bound to its operator. *EMBO J.* **17**, 7404–7415

127. Hernández-Arriaga, A. M., Rubio-Lepe, T. S., Espinosa, M., and del Solar, G. (2009) Repressor CopG prevents access of RNA polymerase to promoter and actively dissociates open complexes. *Nucleic Acids Res.* **37**, 4799–4811

128. Moncalián, G., and De La Cruz, F. (2004) DNA binding properties of protein TrwA, a possible structural variant of the Arc repressor superfamily. *Biochim. Biophys. Acta - Proteins Proteomics* **1701**, 15–23

129. Moncalián, G., Grandoso, G., Llosa, M., and de la Cruz, F. (1997) oriT-processing and regulatory roles of TrwA protein in plasmid R388 conjugation. *J. Mol. Biol.* **270**, 188–200

130. Nelson, W. C., Howard, M. T., Sherman, J. A., and Matson, S. W. (1995) The traY Gene Product and Integration Host Factor Stimulate Escherichia coli DNA Helicase I-catalyzed Nicking at the F Plasmid oriT * manner in the absence of the traY gene product (TraYp). **270**, 28374–28380

131. Lu, J., Wong, J. J. W., Edwards, R. A., Manchak, J., Frost, L. S., and Glover, J. N. M. (2008) Structural basis of specific TraD-TraM recognition during F plasmid-mediated bacterial conjugation. *Mol. Microbiol.* **70**, 89–99

132. Karl, W., Bamberger, M., and Zechner, E. L. (2001) Transfer Protein TraY of Plasmid R1 Stimulates TraI-Catalyzed oriT Cleavage In Vivo. **183**, 909–914

9. APPENDIX

Table A.1 Oligonucleotides used		
Name	**Sequence(5´-3´)**	**Purpose**
DIRECT	TCTCTATTGCCCACTTAT	Used in the primer extension to determine the position of P_c
INVERSE	TTCTAGTTCTTTTTACAC	Used in the primer extension to determine the position of P_c
MARKER1	TTCTAGTTCTTTTTACAC	Used in the primer extension to determine the position of P_r
MARKER2	ACGGTCTAGCGCTTACAAT	Used in the primer extension to determine the position of P_r
Prom28UP Bam	cgcggatccTATACCACCTCGCAAAATAAACC	Used for the fragment I_c of inter-genic region between gene 28 and *rco*$_{LS20}$
Prom28UP Hind	ccccaagcttTATACCACCTCGCAAAATAAACC	Used for the fragment I_c, I_r, IV_c, IV_r, V_c , V_r, VI_c and VI_r of inter-genic region between gene 28 and *rco*$_{LS20}$
Xre prom_New	ttttaagcttGCACCAGCATCAAGTAACACTTGTTTCAG	Used for the fragment I_{Ar} of the intergenic region between gene 28 and *rco*$_{LS20}$
P28_Δ15	tattaagcttGGGGCAAGTTCACACTAACTTTCACTGTG	Used for the fragment IV_c , VII_c, IV_r, and VII_r of inter-genic region between gene 28 and *rco*$_{LS20}$
P28_Δ16	tattaagcttGGCCCTTATCCTGTTTTTACTAAACTT	Used for the fragment V_c, $VIII_c$, V_r, and $VIII_r$ of alongwith inter-genic region between gene 28 and *rco*$_{LS20}$
P28_Up_New	tattaagcttCCCTGTATACGGTCTAGCGCTTAC	Used for the fragment V_c and V_r of alongwith inter-genic region between gene 28 and *rco*$_{LS20}$
OGR1	tattaagcttCCGGTTAAAAATTTCACTGAAATAC	Used for the fragment VII_c and VII_r of inter-genic region between gene 28 and *rco*$_{LS20}$
OGR2	tattaagcttGTCAGTGAAAAAAATGCAGAATAAGG	Used for the fragment $VIII_c$ and $VIII_r$ of inter-genic region between gene 28 and *rco*$_{LS20}$
OPKS8	gggggtcgacGTCCTTTTTTAATTTCATGTATTC	Amplifying rco$_{LS20}$for cloning in pET28b+
OPKS14N	ggggccatggTGGGCAATAGAGAGCAATTTGATC	Amplifying rco$_{LS20}$ for cloning in pET28b+vector
oGR34	GTTCTTTTTACACAGAAATTGTTTG	Gel retardation assay
oGR35	TTTTGATATAGCTCACAGTGAAAGT	Gel retardation assay
oGR36	CCAAGTTGCAGCATGATTTTGAT	Gel retardation assay
oGR37	CTAGAATTCAACTTTTGTTTTAACC	Gel retardation assay

oGR91	AGCGCTTACAATTTTTCGCGTTTTT	Gel retardation assay
oGR92	TTATTCTGCATTTTTTTCACTGACT	Gel retardation assay
oGR93	GGTTTTTTGTTGTTAATCTCAAACA	Gel retardation assay
oGR94	GACTACATTGTGATAGCACACTTTGA	Gel retardation assay
oGR95	TACCAGTTAATTTAACCGTATGTAT	Gel retardation assay
oGR96	ATTTTTAACCGGTTTTTTTCACTGAG	Gel retardation assay
oGR97	ATTTTTCGCGTTTTTTTTACTTAGTG ACCCCCA	Gel retardation assay
oGR98	TTATTCTGCATGGGGGTCACTGACTC TTTCACTAA	Gel retardation assay
oGR99	TTTTGCTAGCGTAAGGATGGAGGAA TTTTCTTGCG	Gel retardation assay
oGR135	TTTTTACACAGAAATTGTTTGAGAT TAACAAGGATCCAAAAAACCTTATTC TGCA	PCR overlapping to add 5 bps to the inter-genic region between genes rcoLS20 and 28.
oGR136	GGATCCTTGTTAATCTCAAACAATTT CTGTGTAAAAAGAACTAGAATTCAA C	PCR overlapping to add 5 bps to the inter-genic region between genes rcoLS20 and 28.
oGR138	GCCTCTCTTTTAAAGCTAAAATGATG TATG	Circularization assay
oGR20	ttttttctagaTTCTGATAATCTCGCTTTCA TTTCATCGTG	Used for cloning fragment 1 in plasmid pUCTA2501 with oGR21
oGR21	ttttttctagaCCGAAAAAGTGAAAATAAA ATTTA	Used for cloning fragment 1 in plasmid pUCTA2501 with oGR20
oGR22	ttttttctagaCTAAATAATGGTTGAACAT AAATGT	Used for cloning fragment 2 in plasmid pUCTA2501 with oGR20
oGR23	ttttttctagaGTACCAGATTTATTGCTGA ATGCA	Used for cloning fragment 3 in plasmid pUCTA2501 with oGR21
oGR28	ttttttctagaAAAGAGCAATCTCGTCATC GAAGAC	Used for cloning fragment 5 in plasmid pUCTA2501 with oGR20 and for cloning fragment 6 in plasmid pUCTA2501with oGR29
oGR29	ttttttctagaTTGTTAACGCTCCTTTTCA TCGATTT	Used for cloning fragment 4 in plasmid pUCTA2501 with oGR21 and for cloning fragment 6 in plasmid pUCTA2501with oGR28
oGR43	ttttgctagcCAAAGTAATGTGCAGAAAT CGATGA	Used for cloning *genes 56, 57* and *58* in vector pDR110 with oGR60.
oGR47	ttttgctagcAGATGATCAAGAAGGAAA TGACCAA	Used for cloning *gene 57* and *58* together in vector pDR110 with oGR60.
oGR56	GAAATCAAAGTGACATTTTAAAGGG	Used to knockout of the oriT region

	GATCT	of the plasmid pLS20
oGR57	ttttgtcgacTTGAAAGACCTTTGATGTTGAGATCCGGCA	Used to knockout the oriT region of the plasmid pLS20
oGR58	ttttggatccAGTATGAAATGGAACAGAGCCGGTAGGCAA	Used to knockout the oriT region of the plasmid pLS20
oGR59	TGCCGTATGTTTGATACAGTTCTAAATATT	Used to knockout the oriT region of the plasmid pLS20
oGR60	ttttgcatgcTCCTTTAATTTCAGAATTGCCTACC	Used for cloning *gene 56, 57* and *58* in vector pDR110 with oGR43
oGR61	ttttgaattcAAAGAGCAATCTCGTCATCGAAGAC	Used for cloning 382 bp upstream *gene 56* in vector pDG1663
oGR62	ttttggatccTTGTTAACGCTCCTTTTCATCGATT	Used for cloning 382 bp upstream *gene 56* in vector pDG1663
oGR77	tttttctagacATTCGCTATGATATACGCTACGCC	Used for cloning fragments of pTA1040 in the vector pUCTA2501 with oGR78
oGR78	tttttctagaAAACAAACACGAAACTTAATTCATC	Used for cloning fragments of pTA1040 in the vector pUCTA2501 with oGR77
oGR79	tttttctagaTCTTTACTATTGTAAAGTCTTATGT	Used for cloning fragments of pTA1040 in the vector pUCTA2501 with oGR77
oGR114	ttttgcatgcGGCGTACTGCCTGAACGAGAAGCTA	Used for cloning the Spectinomycin gene in pTA1040 with primer oGR115
oGR115	ttttgcatgcTGATGCCTCAAGCTAGAGAGTCGAA	Used for cloning the Spectinomycin gene in pTA1040 with primer oGR114
oGR133	ttttactagtCAAAGTAATGTGCAGAAATCGATGA	Used for cloning for *gene 56* in vector pDR110 with oGR134
oGR134	ttttggatccCGCGCTTTAAACACTGCGGCCAGCT	Used for cloning for *gene 56* in vector pDR110 with oGR133
5´- overhang sequences are indicated in lower case and restriction sites are underlined; Mutations or additions marked in bold and underlined		

Table A.2 Plasmids used		
Plasmids	**Description**	**Reference or source**
pDR110	*B. subtilis amyE* integration vector containing IPTG-inducible Pspank promoter	D. Rudner
pDG1663	*B. subtilis thrC* integration vector containing promoter-less *lacZ* gene which is used for promoter screening	Guérout-Fleury AM et al.,1996, BGSC
pLS20*cat*	Native plasmid pLS20 labelled with Cm resistance cassette in unique SalI site.	Itaya et al., 2006
pTZ57R/T	T/A cloning vector	Fermentas
$pDP_{spank}rco_{LS20}$	rco_{LS20} gene is cloned in integration vector pDR110	Singh et.al. 2013
pET28b(+)	Vector for expressing His tagged heterologous proteins in *E.coli*	*Novagene*, Madison, WI
$pDGP_{c(F\_Ic)}$	Fragment I_c cloned in front of *lacZ* gene of pDG1663	This work
$pDGP_{c(F\_Ic+5bp)}$	Fragment I_{c+5bps} cloned in front of *lacZ* gene of pDG1663	This work
$pDGP_{c(F\_IIc)}$	Fragment II_c cloned in front of *lacZ* gene of pDG1663	This work
$pDGP_{c(F\_IIIc)}$	Fragment III_c cloned in front of *lacZ* gene of pDG1663	This work
$pDGP_{c(F\_IVc)}$	Fragment IV_c cloned in front of *lacZ* gene of pDG1663	This work
$pDGP_{c(F\_Vc)}$	Fragment V_c cloned in front of *lacZ* gene of pDG1663	This work
$pDGP_{c(F\_VIIc)}$	Fragment VII_c cloned in front of *lacZ* gene of pDG1663	This work
$pDGP_{c(F\_VIIIc)}$	Fragment $VIII_c$ cloned in front of *lacZ* gene of pDG1663	This work
$pDGP_{r(F\_Ir)}$	Fragment I_r cloned in front of *lacZ* gene of pDG1663	This work
$pDGP_{r(F\_IAr)}$	Fragment I_{Ar} cloned in front of *lacZ* gene of pDG1663	This work
$pDGP_{r(F\_IIIr)}$	Fragment III_r cloned in front of *lacZ* gene of pDG1663	This work
$pDGP_{r(F\_IVr)}$	Fragment IV_r cloned in front of *lacZ* gene of pDG1663	This work
$pDGP_{r(F\_Vr)}$	Fragment V_r cloned in front of *lacZ* gene of pDG1663	This work
$pDGP_{r(F\_VIIr)}$	Fragment VII_r cloned in front of *lacZ* gene of pDG1663	This work
$pDGP_{r(F\_VIIIr)}$	Fragment $VIII_r$ cloned in front of *lacZ*	This work

	gene of pDG1663	
pRco$_{LS20}$-His	*rco$_{LS20}$* in the pET28b(+) vector	This work
pDP$_{spank}$rco$_{LS20}$-His	*rco*$_{LS20}$-His cloned after the P$_{spank}$ of the vector pDR110	This work
pLS20Δ56-58*kancat*	Genes *56-57-58* were deleted and replaced by kanamycin gene	This work
pUCTA2501	Em containing vector of pTA1015	Thesis Meijer
pAXO1	plasmid allowing the integration of DNA fragments at the *B. subtilis lacA* locus	J. Bacteriol. 2001;183:2696-2699
pBEST501	*E. coli* vector containing neomycin resistance marker in multiple cloning site	Nucleic Acids Res 17: 4410.
pGR8A	Fragment 1 cloned in vector pUCTA2501, orientation A	This work
pGR8B	Fragment 1 cloned in vector pUCTA2501, orientation B	This work
pGR10A	Fragment 2 cloned in vector pUCTA2501, orientation A	This work
pGR10B	Fragment 2 cloned in vector pUCTA2501, orientation B	This work
pGR12A	Fragment 3 cloned in vector pUCTA2501, orientation A	This work
pGR12B	Fragment 3 cloned in vector pUCTA2501, orientation B	This work
pGR16A	Fragment 6 cloned in vector pUCTA2501, orientation A	This work
pGR16B	Fragment 6 cloned in vector pUCTA2501, orientation B	This work
pGR20A	Fragment 4 cloned in vector pUCTA2501, orientation A	This work
pGR20B	Fragment 4 cloned in vector pUCTA2501, orientation B	This work
pGR22A	Fragment 5 cloned in vector pUCTA2501, orientation A	This work
pGR22B	Fragment 5 cloned in vector pUCTA2501, orientation B	This work
pGR42	Genes *57* and *58* cloned in pDR110	This work
pGR43	Gene *56* cloned in pAXO1	This work
pGR52	Genes *56-57 cloned in pDR110*	This work

Table A.3: Strains used		
Strains	**Genotype or description**	**References or source**
Escherichia coli		
XL1-Blue	endA1 gyrA96(nal^R) thi-1 recA1 relA1 lac glnV44 F'[::Tn10 proAB$^+$ lacI^q Δ(*lacZ*)M15] hsdR17(r_K^- m_K^+)	Bullock et al., 1987
Bacillus subtilis		
168 (1A700)	*trpC2*	BGSC*
PKS3	*trpC2 thrC*::$P_{c(F\_Ic)}$ -*lacZ* (*em*)	This work
PKS5	*trpC2 thrC*::$P_{c(F\_Ic)}$ -*lacZ*(*em*) *amyE*::P_{spank}*rco*$_{LS20}$ (*spc*)	This work
PKS7	*trpC2 thrC*::Δ -*lacZ*(*em*)	This work
PKS8	*trpC2 thrC*::$P_{c(F\_Ic)}$-*lacZ*(*em*) containing plasmid pLS20*cat*	This work
PKS11	*trpC2* containing plasmid pLS20*cat*	Singh P.K. et.al.,2012
PKS30	*trpC2 thrC*::$P_{c(F\_IIIc)}$-*lacZ*(*em*)	This work
PKS31	*trpC2 thrC*::$P_{r(F\_IIIr)}$-*lacZ*(*em*)	This work
PKS32	*trpC2 thrC*::$P_{c(F\_IIIc)}$-*lacZ*(*em*), containing plasmid pLS20*cat*	This work
GR9	*trpC2 thrC*::$P_{r(F\_IIIr)}$-*lacZ*(*em*), containing plasmid pLS20*cat*	This work
GR10	*trpC2 thrC*::$P_{c(F\_IIc)}$-*lacZ*(*em*)	This work
GR11	*trpC2 thrC*::$P_{c(F\_IIc)}$-*lacZ*(*em*), containing plasmid pLS20*cat*	This work
GR12	*trpC2 thrC*::$P_{c(F\_IIc)}$-*lacZ*(*em*) *amyE*::P_{spank}*rco*$_{LS20}$ (*spc*)	This work
GR14	*trpC2 thrC*::$P_{r(F\_IIIr)}$-*lacZ*(*em*) *amyE*::P_{spank}*rco*$_{LS20}$ (*spc*)	This work
GR16	*trpC2 thrC*::$P_{c(F\_IIIc)}$-*lacZ*(*em*) amy*E*::P_{spank}*rco*$_{LS20}$ (*spc*)	This work
GR25	*trpC2 thrC*::$P_{r(F\_Ir)}$-*lacZ* (*em*)	This work
GR27	*trpC2 thrC*::$P_{r(F\_IVr)}$-*lacZ* (*em*)	This work
GR28	*trpC2 thrC*::$P_{c(F\_IVc)}$-*lacZ* (*em*)	This work
GR29	*trpC2 thrC*::$P_{r(F\_Vr)}$-*lacZ (em*)	This work
GR30	*trpC2 thrC*::$P_{c(F\_Vc)}$-*lacZ* (*em*)	This work
GR33	*trpC2 thrC*::$P_{r(F\_IVr)}$-*lacZ* (*em*), containing plasmid pLS20*cat*	This work
GR34	*trpC2 thrC*::$P_{c(F\_IVc)}$-*lacZ* (*em*), containing plasmid pLS20*cat*	This work
GR35	*trpC2 thrC*::$P_{r(F\_Vr)}$-*lacZ (em*), containing plasmid pLS20*cat*	This work
GR36	*trpC2 thrC*::$P_{c(F\_Vc)}$*lacZ* (*em*), containing plasmid pLS20*cat*	This work
GR39	*trpC2 thrC*::$P_{r(F\_Ir)}$*lacZ* (*em*), containing plasmid pLS20*cat*	This work
GR42	*trpC2 thrC*::$P_{c(F\_IVc)}$*lacZ* (*em*) *amyE*::P_{spank}*rco*$_{LS20}$ (*spc*)	This work

GR43	*trpC2 thrC*::$P_{c(F\_Vc)}$*lacZ* (*em*) *amyE*::P_{spank}*rco*$_{LS20}$ (*spc*)	This work
GR62	*trpC2 thrC*::$P_{r(F\_IAr)}$*lacZ* (*em*)	This work
GR66	*trpC2 thrC*::$P_{r(F\_IAr)}$*lacZ* (*em*), containing plasmid pLS20*cat*	This work
GR68	*trpC2 thrC*::$P_{c(F\_VIIc)}$*lacZ* (*em*)	This work
GR69	*trpC2 thrC*::$P_{r(F\_VIIr)}$*lacZ* (*em*)	This work
GR70	*trpC2 thrC*::$P_{c(F\_VIIIc)}$*lacZ* (*em*)	This work
GR71	*trpC2 thrC*::$P_{r(F\_VIIIr)}$*lacZ* (*em*)	This work
GR76	*trpC2 thrC*::$P_{r(F\_IAr)}$*lacZ* (*em*) amy*E*::P_{spank}*rco*$_{LS20}$ (*spc*)	This work
GR82	*trpC2 thrC*::$P_{r(F\_VIIr)}$*lacZ* (*em*), containing plasmid pLS20*cat*	This work
GR83	*trpC2 thrC*::Δ *lacZ*(*em*), containing plasmid pLS20*cat*	This work
GR90	trpC2 *amyE*::P_{spank}rco$_{LS20}$-his	This work
GR92	*trpC2 thrC*::$P_{r(F\_Ir)}$-*lacZ* (*em*) *amyE*::P_{spank}*rco*$_{LS20}$ (*spc*)	This work
GR97	*trpC2 thrC*::$P_{r(F\_IVr)}$-*lacZ* (*em*) *amyE*::P_{spank}*rco*$_{LS20}$ (*spc*)	This work
GR102	*trpC2 thrC*::$P_{r(F\_Vr)}$-*lacZ* (*em*) *amyE*::P_{spank}*rco*$_{LS20}$ (*spc*)	This work
GR116	*trpC2 thrC*::$P_{r(F\_VIIIr)}$-*lacZ* (*em*), containing plasmid pLS20*cat*	This work
GR164	*trpC2 thrC*::$P_{c(F\_VIIc)}$-*lacZ* (*em*) amy*E*::P_{spank}*rco*$_{LS20}$ (*spc*)	This work
GR165	*trpC2 thrC*::$P_{c(F\_VIIc)}$-*lacZ* (*em*), containing plasmid pLS20*cat*	This work
GR189	*trpC2 thrC*::$P_{c(F\_Ic+5)}$-*lacZ* (*em*)	This work
GR191	*trpC2 thrC*::$P_{c(F\_Ic+5)}$-*lacZ* (*em*), containing plasmid pLS20*cat*	This work
GR195	*trpC2 thrC*::$P_{c(F\_Ic+5)}$-*lacZ* (*em*) *amyE*::P_{spank}*rco*$_{LS20}$ (*spc*)	This work
GR81	*trp*C2, pUCTA2501 (*em*)	This work
GR104	*trpC2*, pUCTA2501(*em*), pLS20*cat*	This work
GR114	*trpC2,* pGR10(*em*),pLS20*cat*	This work
GR115	*trpC2,*pGR12(*em*),pLS20cat	This work
GR121	*trpC2,* pGR11(*em*), pLS20*cat*	This work
GR122	*trpC2,* pGR14(*em*), pLS20*cat*	This work
GR124	*trpC2,* pGR8(*em*), pLS20*cat*	This work
GR127	*trpC2, amy*E::P_{spank}58$_{LS20}$,	This work
GR137	*trpC2,* pGR15(*em*), pLS20*cat*	This work

GR138	*trpC2,* pGR16(*em*), pLS20*cat*	This work
GR139	*trpC2,* pGR21(*em*), pLS20*cat*	This work
GR140	*trpC2,* pGR29(*em*), pLS20*cat*	This work
GR149	*trpC2,* pLS20Δ56-58*kancat*	This work
GR150	*trpC2*, *amyE*::P_{spank}*565758spec* pLS20Δ565758*kancat*	This work
GR153	*trpC2, amyE*::P_{spank}*58spec* pLS20Δ565758*cat*	This work
GR183	*trpC2,* pGR9(*em*), pLS20*cat*	This work
GR184	*trpC2,* pGR20(*em*), pLS20*cat*	This work
GR185	*trpC2,* pGR22(*em*), pLS20*cat*	This work
GR193	*trpC2, amy*E::$P_{spank}58_{LS20}$, *lac*A::$P_{xyl}56_{LS20}$	This work
GR197	*trpC2, amy*E::P_{spank} 5758_{LS20} (*spec*) pLS20Δ56-58*kancat*	This work
GR200	*trpC2, amy*E::$P_{spank}58_{LS20}$, *lac*A::$P_{xyl}56_{LS20}$, pLS20Δ56-58*kancat*	This work
GR206	*trpC2, amy*E::P_{spank} 5657_{LS20} (*spec*) pLS20Δ56-58*kancat*	This work

*, BGSC: Bacillus Genetic Stock Center, Department of Biochemistry, The Ohio State University, Columbus, OH, USA. (http://www.bgsc.org/)

INSTITUT PASTEUR

Research in Microbiology 161 (2010) 772–782

www.elsevier.com/locate/resmic

Complete nucleotide sequence and determination of the replication region of the sporulation inhibiting plasmid p576 from *Bacillus pumilus* NRS576

Praveen K. Singh[1], Sandra Ballestero-Beltrán[1], Gayetri Ramachandran, Wilfried J.J. Meijer*

Centro de Biología Molecular "Severo Ochoa" (CSIC-UAM), Instituto de Biología Molecular "Eladio Viñuela" (CSIC), Universidad Autónoma, Canto Blanco, 28049 Madrid, Spain

Received 30 June 2010; accepted 27 July 2010
Available online 21 September 2010

Abstract

Large plasmids, presumably replicating via the theta mechanism, have been identified in numerous gram-positive bacteria. However, their characterization is rather poor and predominantly limited to those harbored by some (opportunistic) pathogenic bacteria. Here we determined the DNA sequence of the 43.3 kb plasmid p576 from *Bacillus pumilus* strain NRS576, the first *B. pumilus* theta-replicating plasmid sequenced. Plasmid p576 has a modular structure, but surprisingly, it does not seem to encode a Rep protein found on most theta-replicating plasmids. However, a ~1 kb region was identified showing homology with the Rep-independent replication region of *Bacillus subtilis* plasmid pLS20, and we demonstrated that this region is sufficient for autonomous replication. The plasmid contains various large direct repeat sequences. A likely function could be attributed to at least 15 putative p576 genes. Some of these are predicted to be involved in stable maintenance of the plasmid; others are likely to encode proteins involved in conjugation. p576 also carries a *rap–phr* cassette whose possible function is discussed.

Keywords: *Bacillus pumilus*; Plasmid; Conjugation; DNA replication; Sporulation

1. Introduction

Circular plasmids are commonly present in gram-positive and -negative bacteria. Based on their mode of replication, they can be grouped into two classes. One group replicates via the rolling-circle mechanism and the other via the theta mechanism. Generally, small (<10 kb) and large plasmids replicate via the rolling circle and theta mechanism of replication, respectively (for review see, Meijer et al., 1998). Various large theta-replicating plasmids from gram-negative bacteria have been studied extensively. However, although numerous large plasmids have been identified in gram-positive bacteria, they are rather poorly characterized, with a few exceptions (Titok et al., 2003; Lioy et al., 2010). Many large plasmids carry conjugation genes permitting them to be transferred horizontally, sometimes even to distantly related species. In addition, rolling circle replicating plasmids, which on their own cannot be transferred horizontally, can often be transferred (via the process termed mobilization) when co-resident with a large conjugative theta-type plasmid. Thus, together with bacteriophages and transposons, conjugative plasmids form part of a large prokaryotic mobile gene pool or "mobilome" that, in combination with (homologous) recombination, can enable rapid and profound prokaryotic genomic rearrangements (for review see, Frost et al., 2005; Gogarten and Townsend, 2005). This warrants a better understanding of gram-positive theta-replicating plasmids. Our attention was caught by a publication describing the presence of a naturally occurring 45 kb plasmid, p576, in the *Bacillus pumilus* strain NRS576 (Lovett and Bramucci, 1974). No complete or partial DNA sequence of any *B. pumilus* theta-replicating plasmid is available. In addition, the presence of p576 appears to inhibit sporulation of its host (Lovett, 1973). As a first impetus to improving our knowledge of this gram-positive plasmid, we

* Corresponding author. Tel.: +34 91 196 4515; fax: +34 91 196 4420.
E-mail addresses: psingh@cbm.uam.es (P.K. Singh), sballestero@cbm.uam.es (S. Ballestero-Beltrán), gramachandran@cbm.uam.es (G. Ramachandran), wmeijer@cbm.uam.es (W.J.J. Meijer).
[1] Contributed equally.

0923-2508/$ - see front matter © 2010 Institut Pasteur. Published by Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.resmic.2010.07.007